

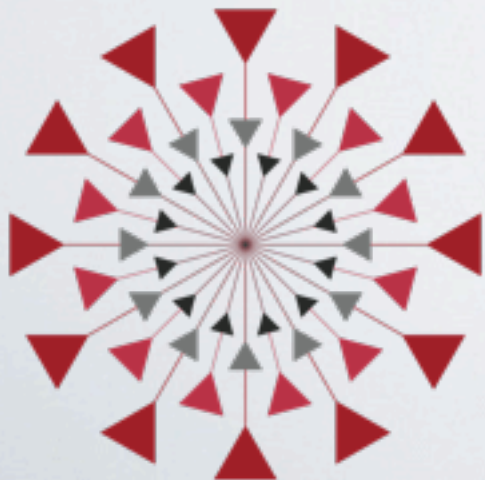
DISCOVERING IMPLICIT NETWORKS FROM POINT PROCESS DATA

Ryan P. Adams

School of Engineering and Applied Sciences
Harvard University

Joint work with Scott Linderman

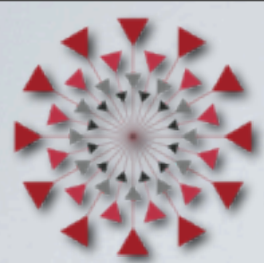
<http://hips.seas.harvard.edu>



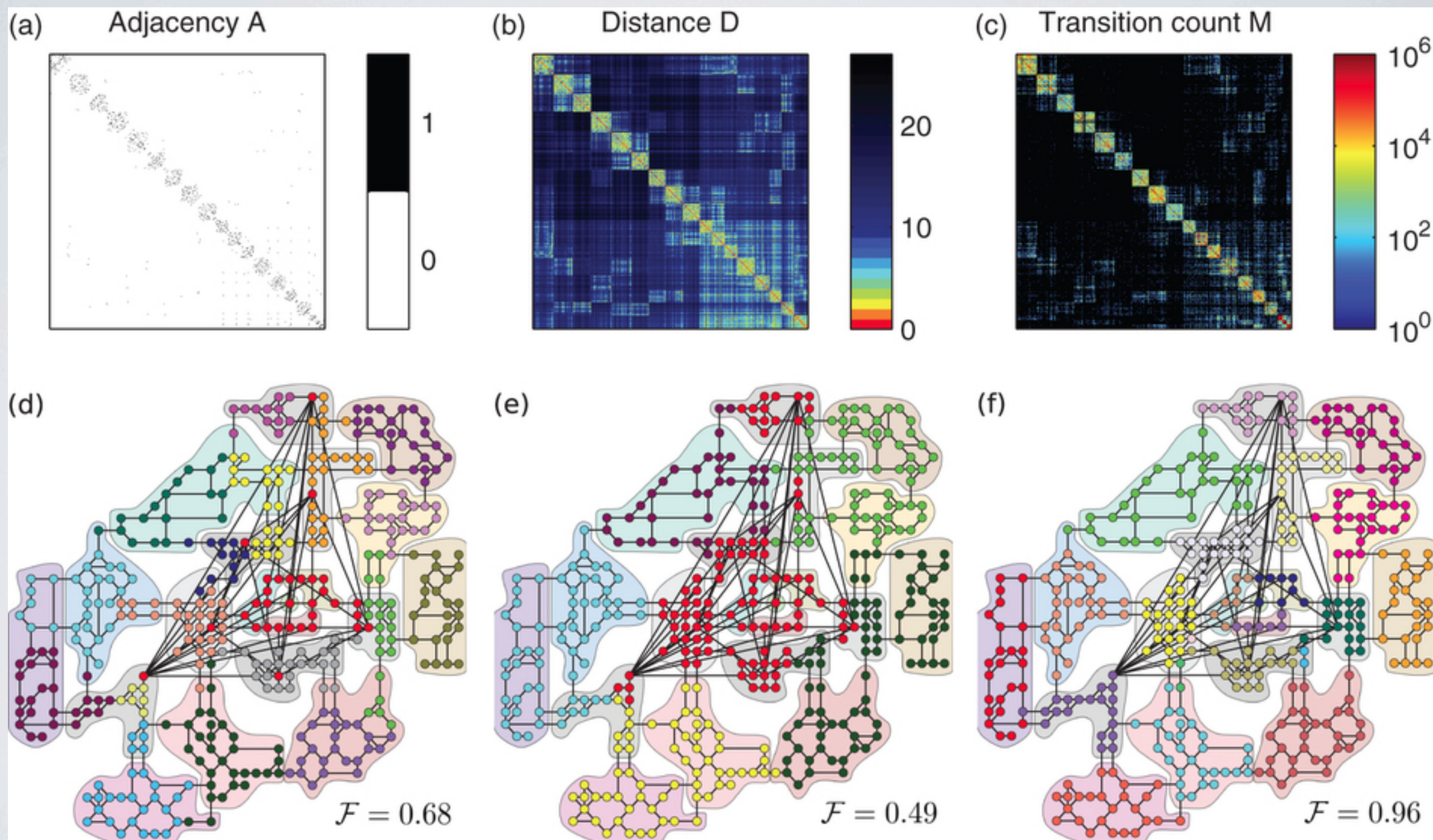
HARVARD
**INTELLIGENT
PROBABILISTIC
SYSTEMS**



Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 03 AUG 2013		2. REPORT TYPE		3. DATES COVERED 00-00-2013 to 00-00-2013	
4. TITLE AND SUBTITLE Discovering Implicit Networks from Point Process Data				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Harvard University ,School of Engineering and Applied Sciences,Cambridge,MA,02138				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES Third Biannual Duke University Workshop on Sensing and Analysis of High-Dimensional Data, July 23-25, 2013, Durham, NC. U.S. Government or Federal Rights License					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 59	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

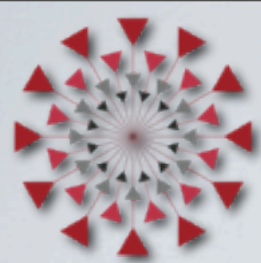


SOCIAL NETWORK ANALYSIS



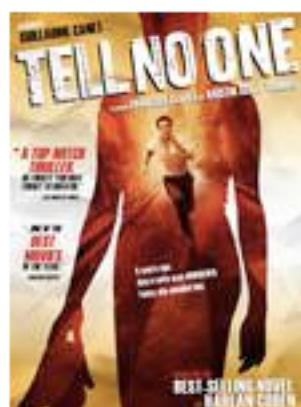
Szell et al, Nature 2012





COLLABORATIVE FILTERING

FOREIGN SUGGESTIONS (about 104) [See all >](#)



Tell No One

Because you enjoyed:
[Memento](#)
[Syriana](#)
[Children of Men](#)

Add



☐ Not Interested



Let the Right One In

Because you enjoyed:
[Seven Samurai](#)
[This Is Spinal Tap](#)
[The Big Lebowski](#)

Add



☐ Not Interested



I've Loved You So Long

Because you enjoyed:
[The Queen](#)
[Syriana](#)
[Good Night, and Good Luck](#)

Add



☐ Not Interested



Downfall

Because you enjoyed:
[Das Boot](#)
[The Killing Fields](#)
[Seven Samurai](#)

Add



☐ Not Interested

DRAMA SUGGESTIONS (about 82) [See all >](#)



The Wrestler

Because you enjoyed:
[Sin City](#)
[Reservoir Dogs](#)
[The Big Lebowski](#)

Add



☐ Not Interested



The Visitor

Because you enjoyed:
[Gandhi](#)
[The Motorcycle Diaries](#)
[The Queen](#)

Add



☐ Not Interested



Brick

Because you enjoyed:
[The Big Lebowski](#)
[Rushmore](#)
[Fight Club](#)

Add



☐ Not Interested



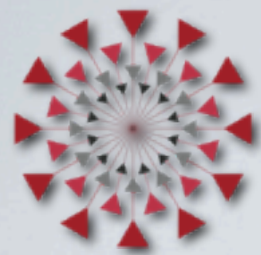
The Pianist

Because you enjoyed:
[Amadeus](#)
[The Killing Fields](#)
[Empire of the Sun](#)

Add

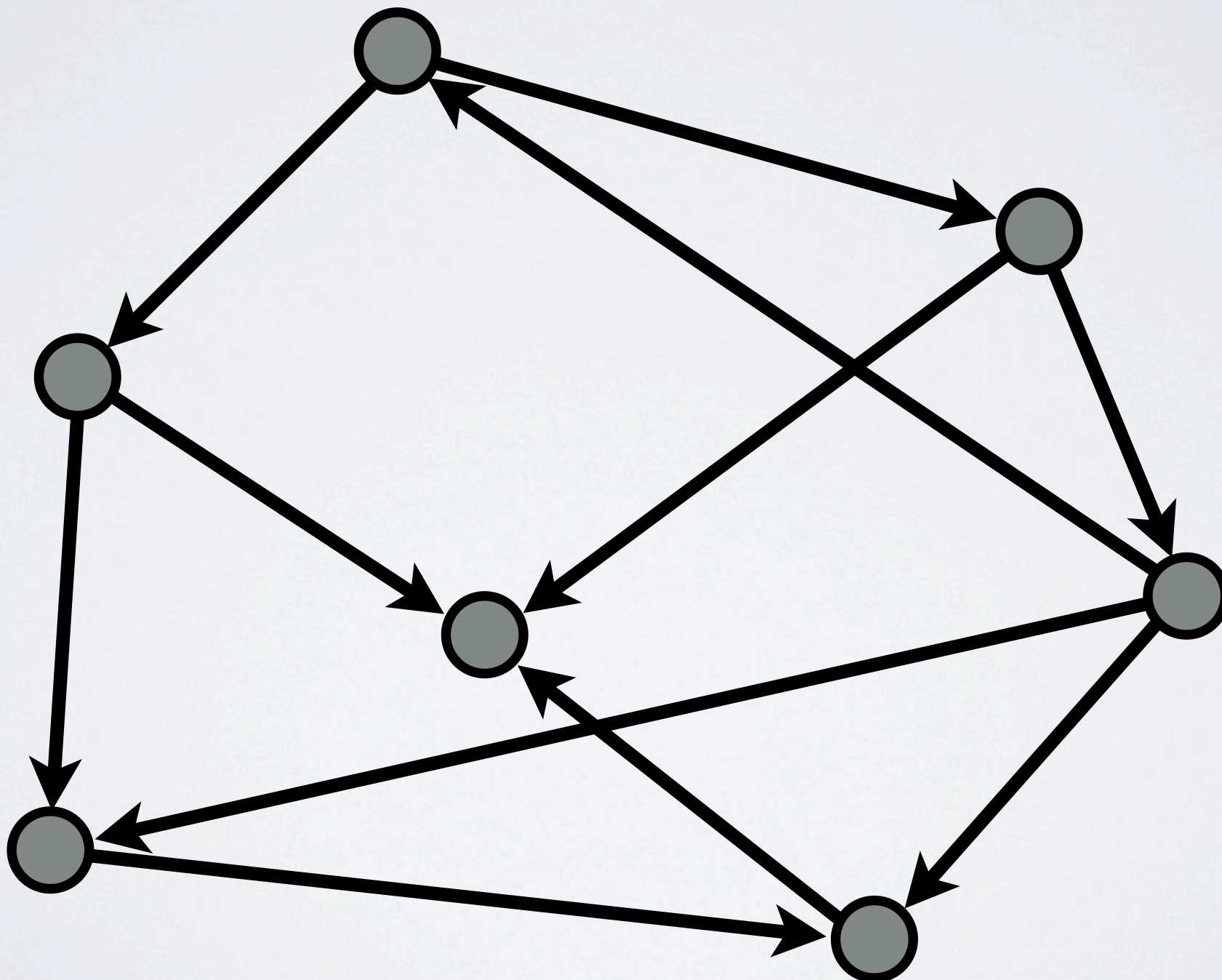


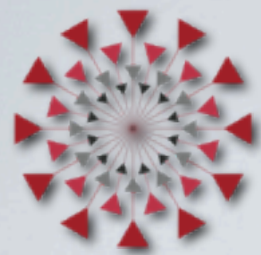
☐ Not Interested



EXPLICIT VS. IMPLICIT NETWORKS

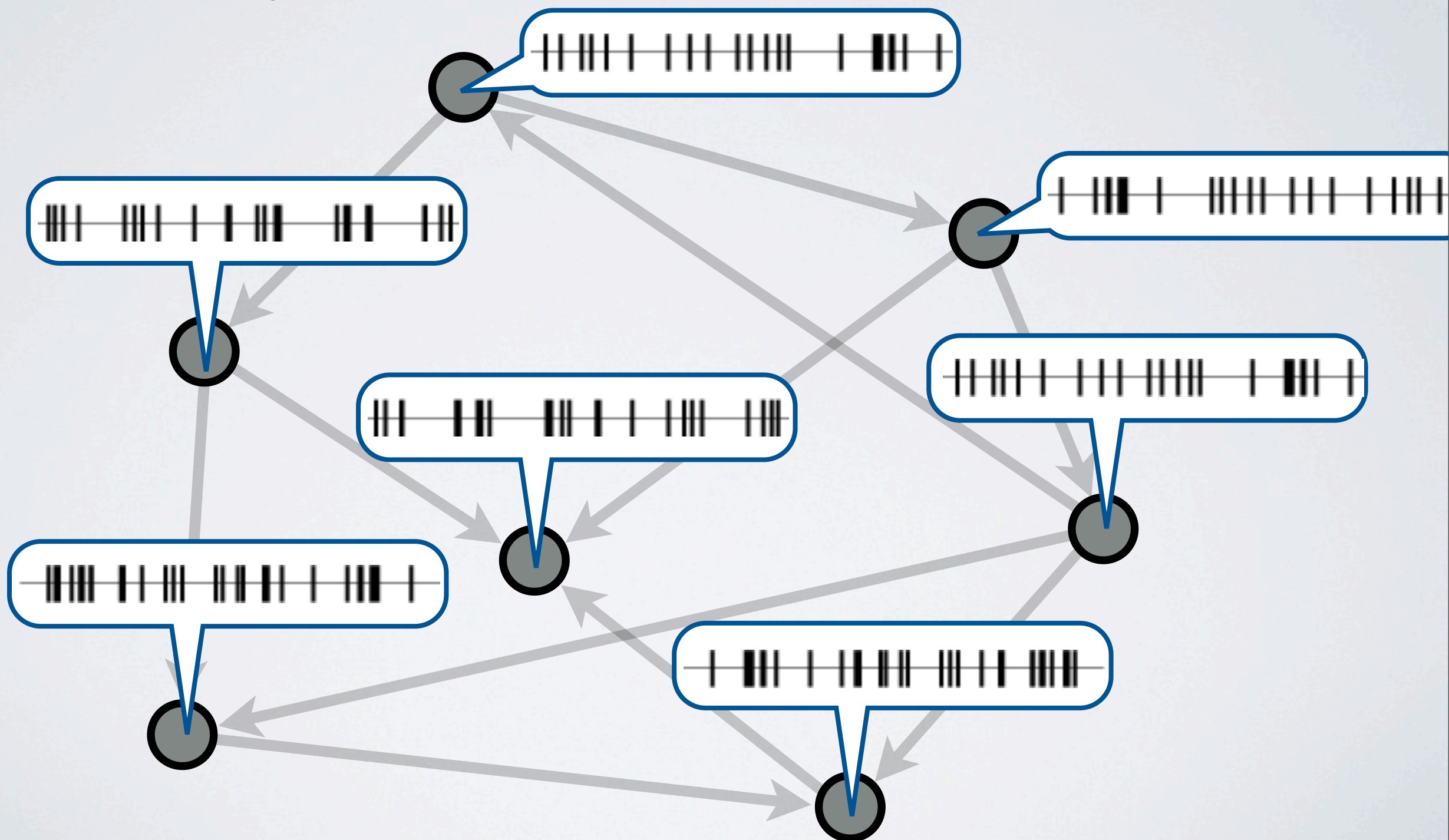
Typically, we see edges and reason about the latent properties of the vertices.

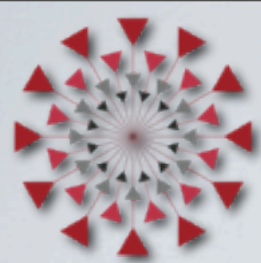




EXPLICIT VS. IMPLICIT NETWORKS

What if we don't observe edges, but only noisy emissions from each vertex?





FUNCTIONAL CONNECTIVITY OF NEURONS

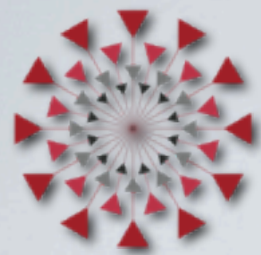
Ensemble raster



Truccolo, Hochberg & Donoghue, 2010



The map displays the spatial distribution of Battery (black dots) and Murder (red crosses) incidents in the Chicago area. The x-axis represents Longitude, ranging from -87.9 to -87.5. The y-axis represents Latitude, ranging from 41.6 to 42.1. The map includes labels for various cities and parks, such as Wilmette, Glenview, Morton Grove, Skokie, Evanston, Park Ridge, Niles, Lincolnwood, Schiller Park, Norridge, Elmwood Park, Melrose Park, River Forest, Bellwood, Maywood, Forest Park, Oak Park, Cicero, Berwyn, La Grange Park, Brookfield, La Grange, Lyons, Summit, Bridgeview, Justice, Burbank, Hickory Hills, Oak Lawn, Chicago Ridge, Palos Hills, Worth, Palos Heights, Alsip, Crestwood, Midlothian, Oak Forest, Harvey, South, Delton, Riverdale, and Calumet City. Major highways are also labeled, including I 294, I 90, I 541, I 55, I 54, I 94, I 90, and IN 912. The legend indicates that black dots represent Battery incidents and red crosses represent Murder incidents. The map shows a high density of both types of incidents in the central and eastern parts of the city, particularly in the areas around the Loop and the South Loop.



OVERVIEW

- ▶ Mutually-Exciting Point Processes
- ▶ Aldous-Hoover Graph Priors
- ▶ MCMC Inference with Data Augmentation
- ▶ Application Examples
- ▶ Extending for Neural Models with Inhibition



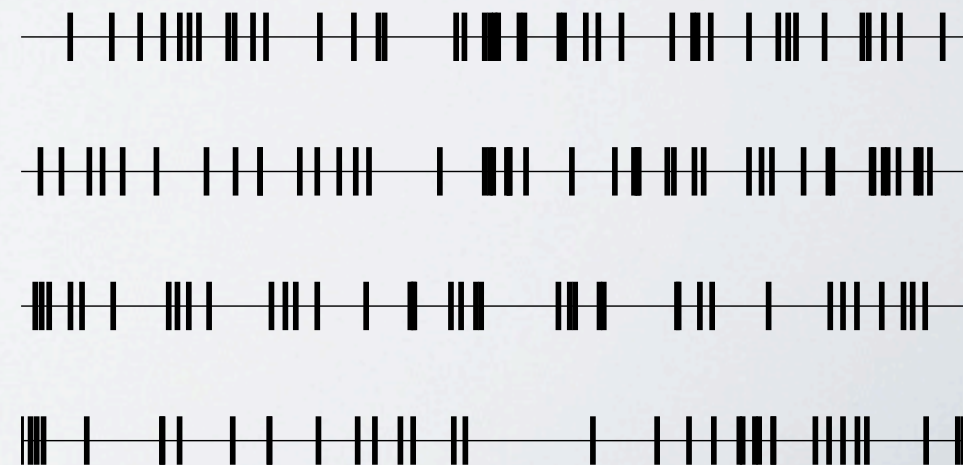
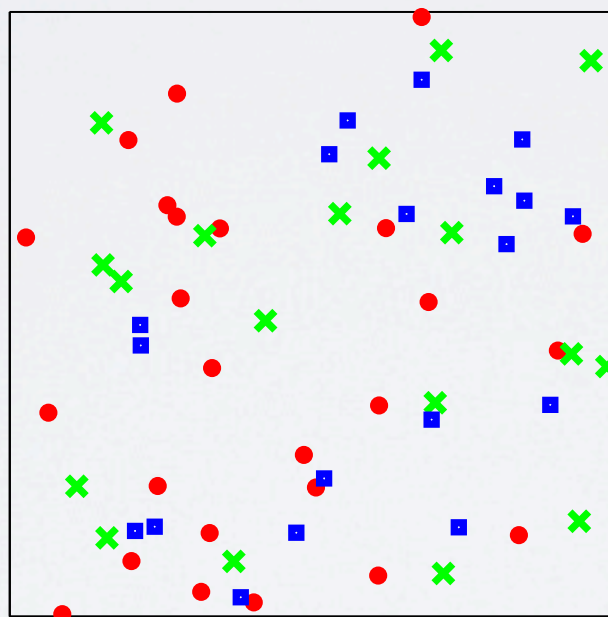
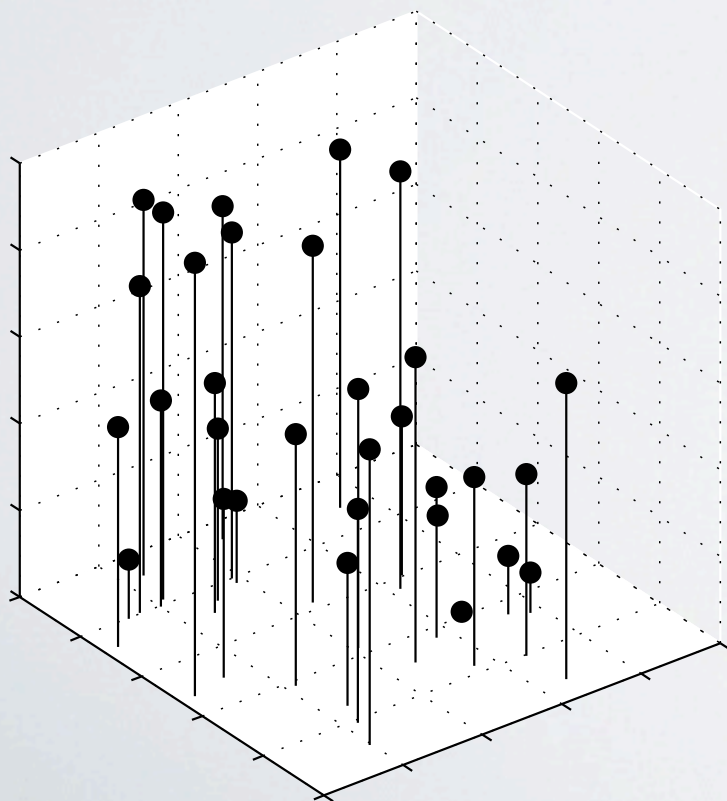
MULTIVARIATE POINT PROCESSES

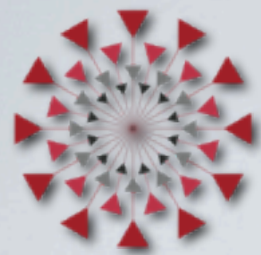
- ▶ The point process is a foundational statistical object.
- ▶ Gives us random subsets of a larger space.
- ▶ Many data are well modeled as point processes:
 - ▶ Seismology
 - ▶ Epidemiology
 - ▶ Economics
- ▶ Modeling dependence is challenging - “beyond Poisson”
 - ▶ Strauss and Gibbs Processes
 - ▶ Determinantal and Permanent Point Processes
 - ▶ TODAY: Mutually-Exciting (Hawkes) Processes



POINT PROCESS

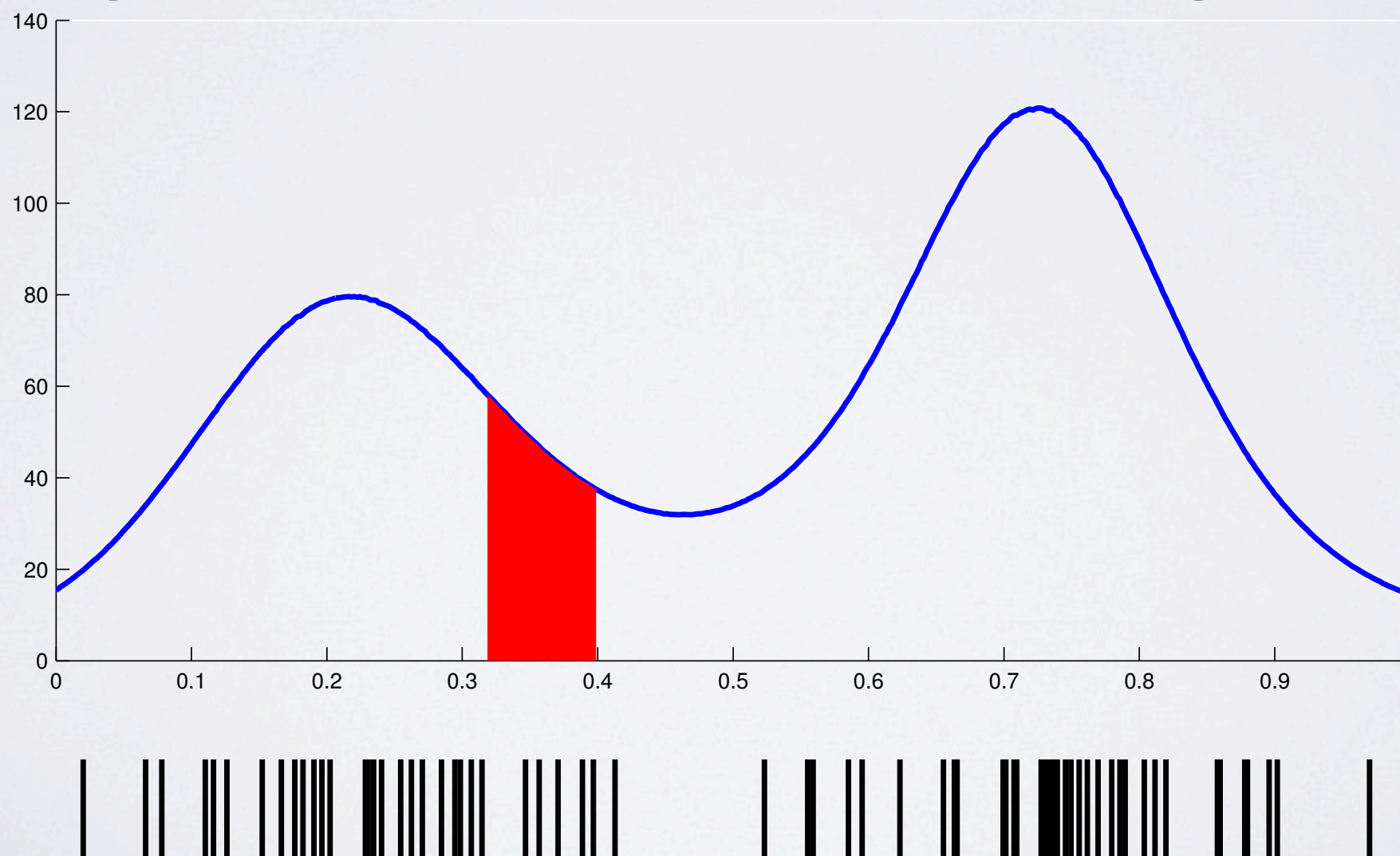
- ▶ A point process on \mathcal{X} gives us random subsets $\{x_n\}_{n=1}^N$.
- ▶ Formally, a random locally-finite counting measure.
- ▶ Most of the time we think of them as giving us finite subsets of compact subset of \mathbb{R}^d , e.g., time or space.





POISSON PROCESS

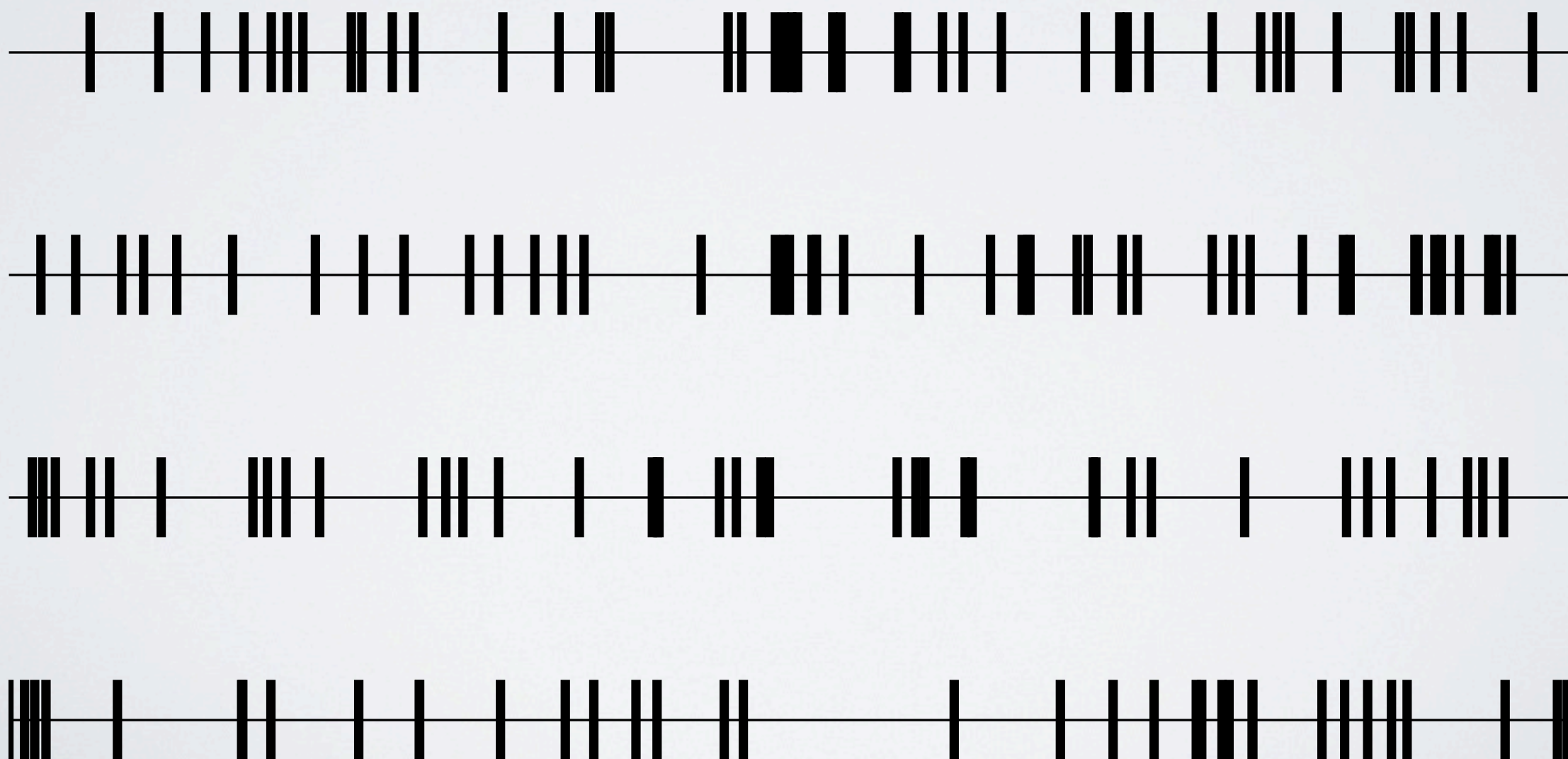
- ▶ The Poisson process is the most basic point process.
- ▶ Disjoint regions are independent.
- ▶ The number of points in a region is determined by integrating the rate function over that region.





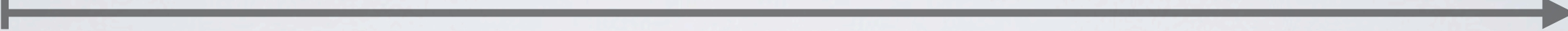
DEPENDENCE IN MULTIVARIATE POINT PROCESSES

- ▶ In the Poisson process, everything is conditionally independent given the rate function.
- ▶ How can we get spike-driven dynamics?

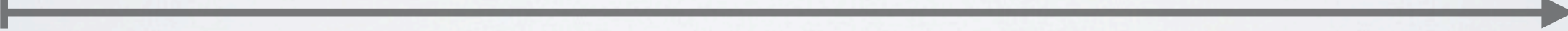


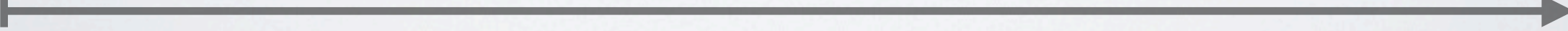


HAWKES PROCESS DYNAMICS

I 

II 

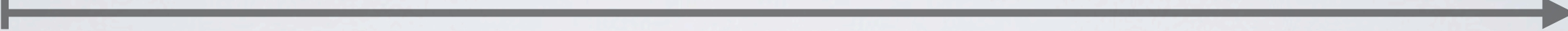
III 

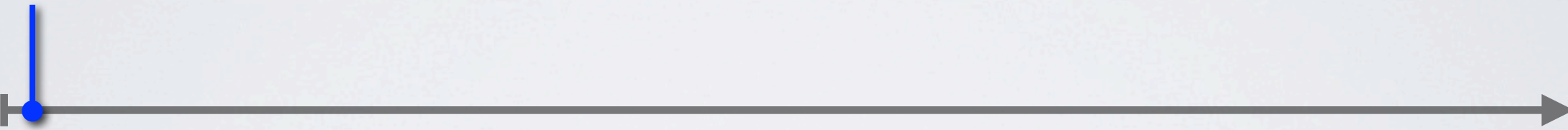
IV 

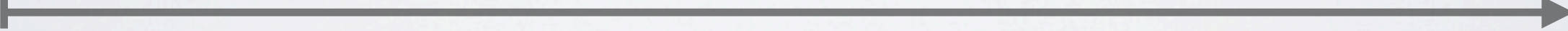
time

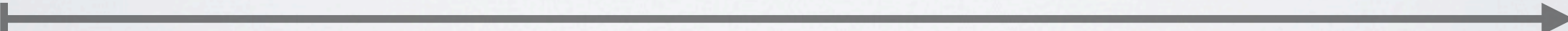


HAWKES PROCESS DYNAMICS

I 

II 

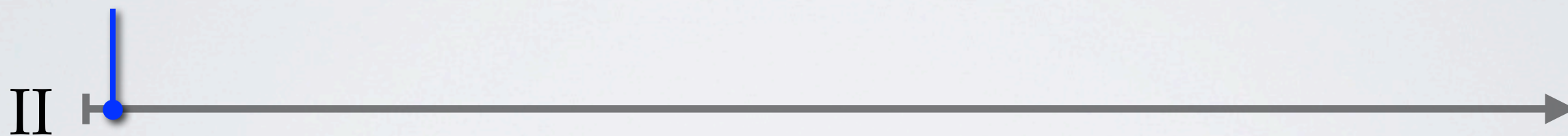
III 

IV 

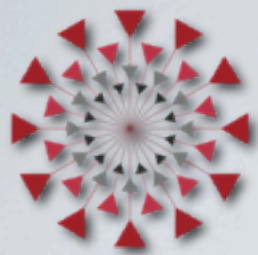
time



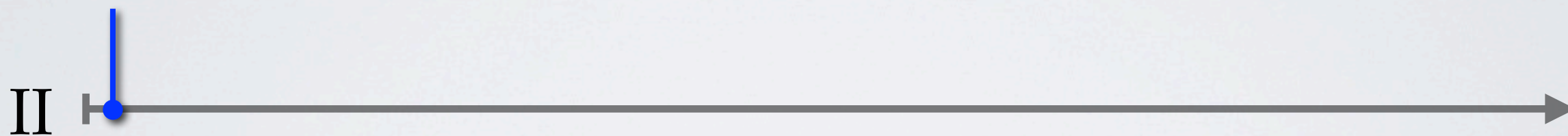
HAWKES PROCESS DYNAMICS



time



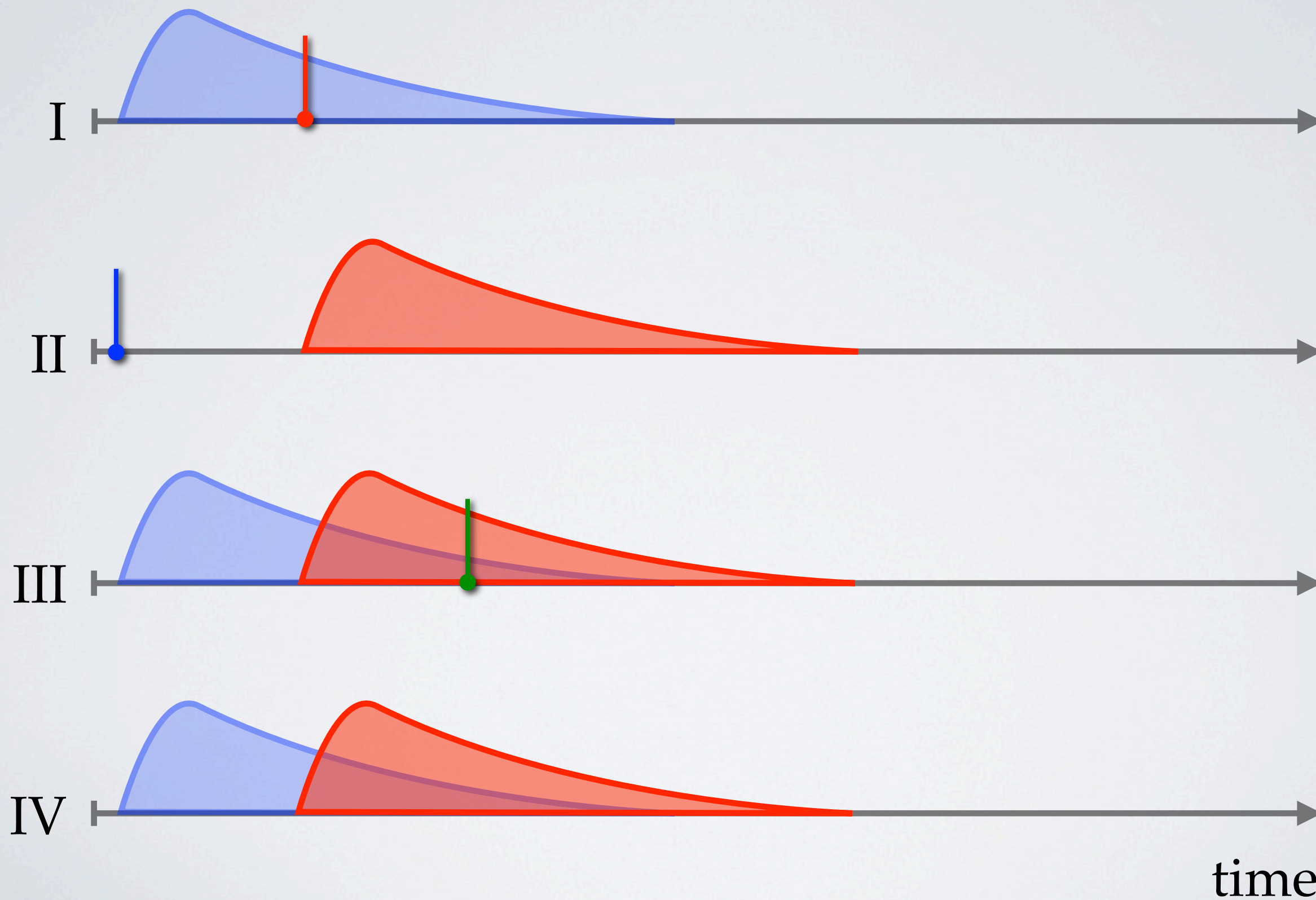
HAWKES PROCESS DYNAMICS

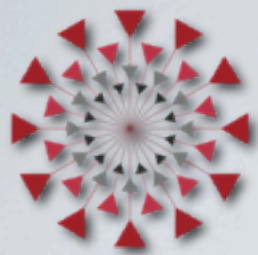


time

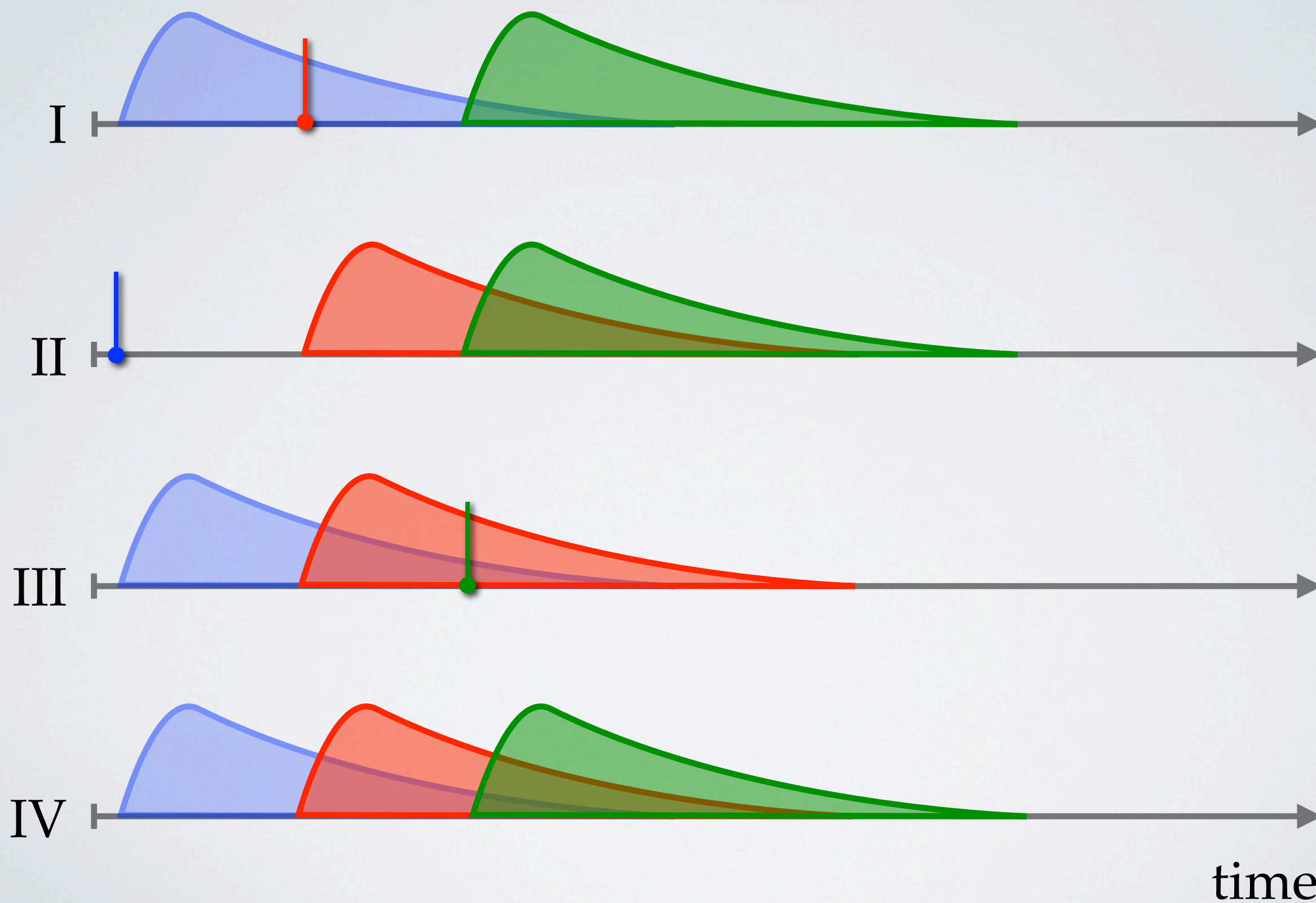


HAWKES PROCESS DYNAMICS



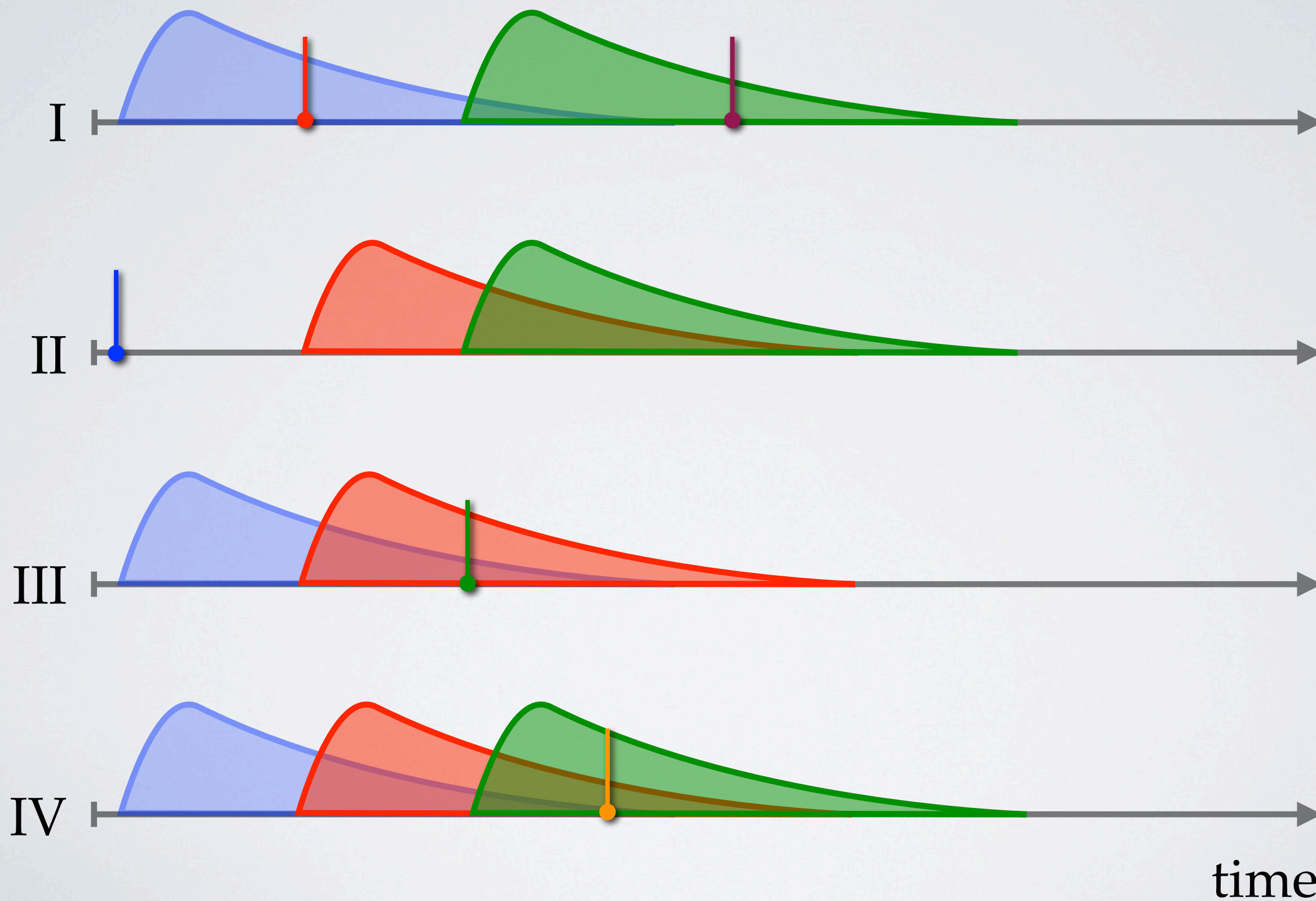


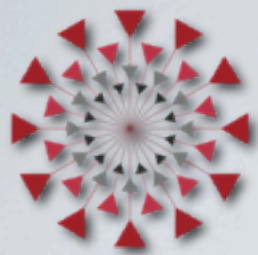
HAWKES PROCESS DYNAMICS



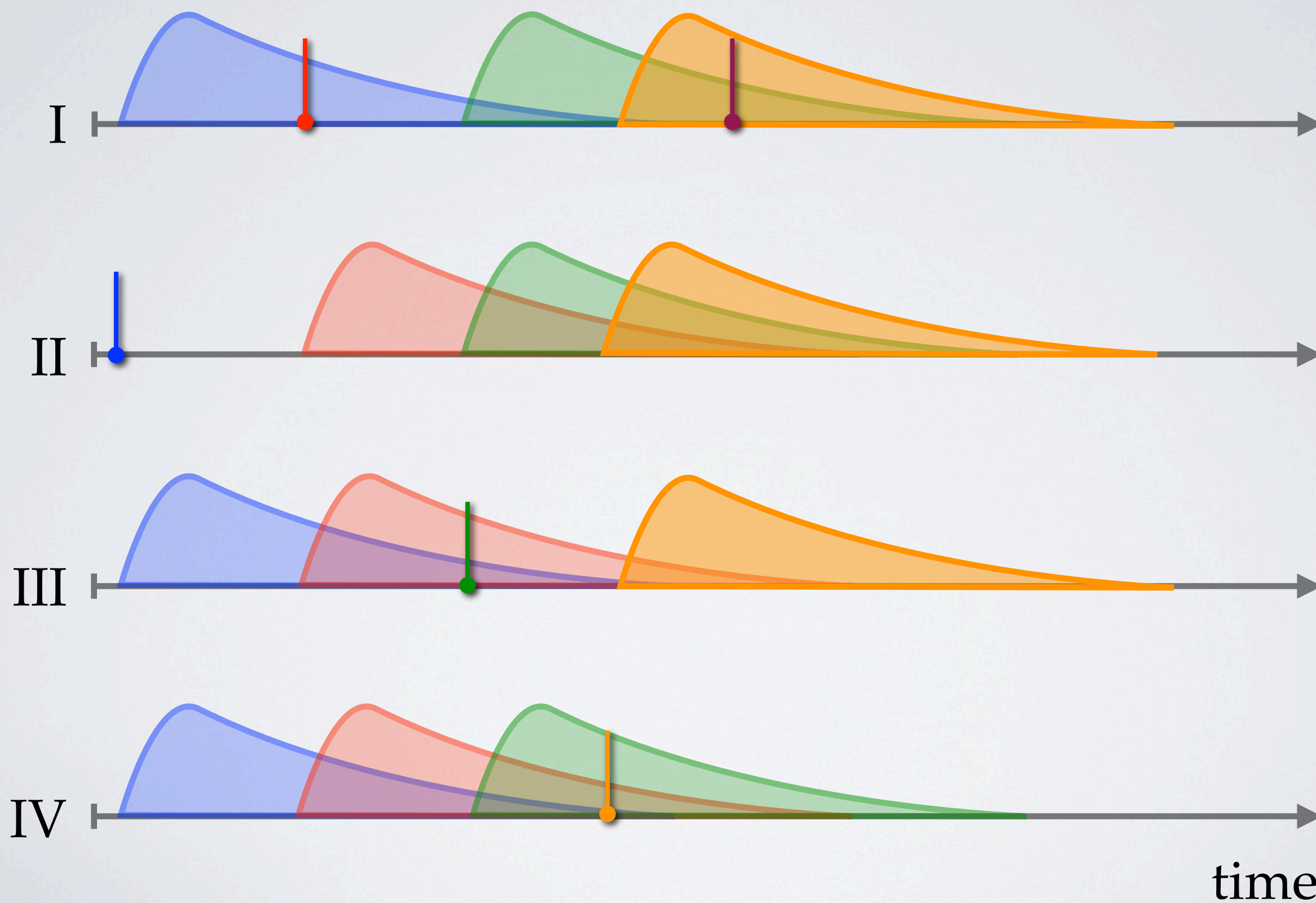


HAWKES PROCESS DYNAMICS



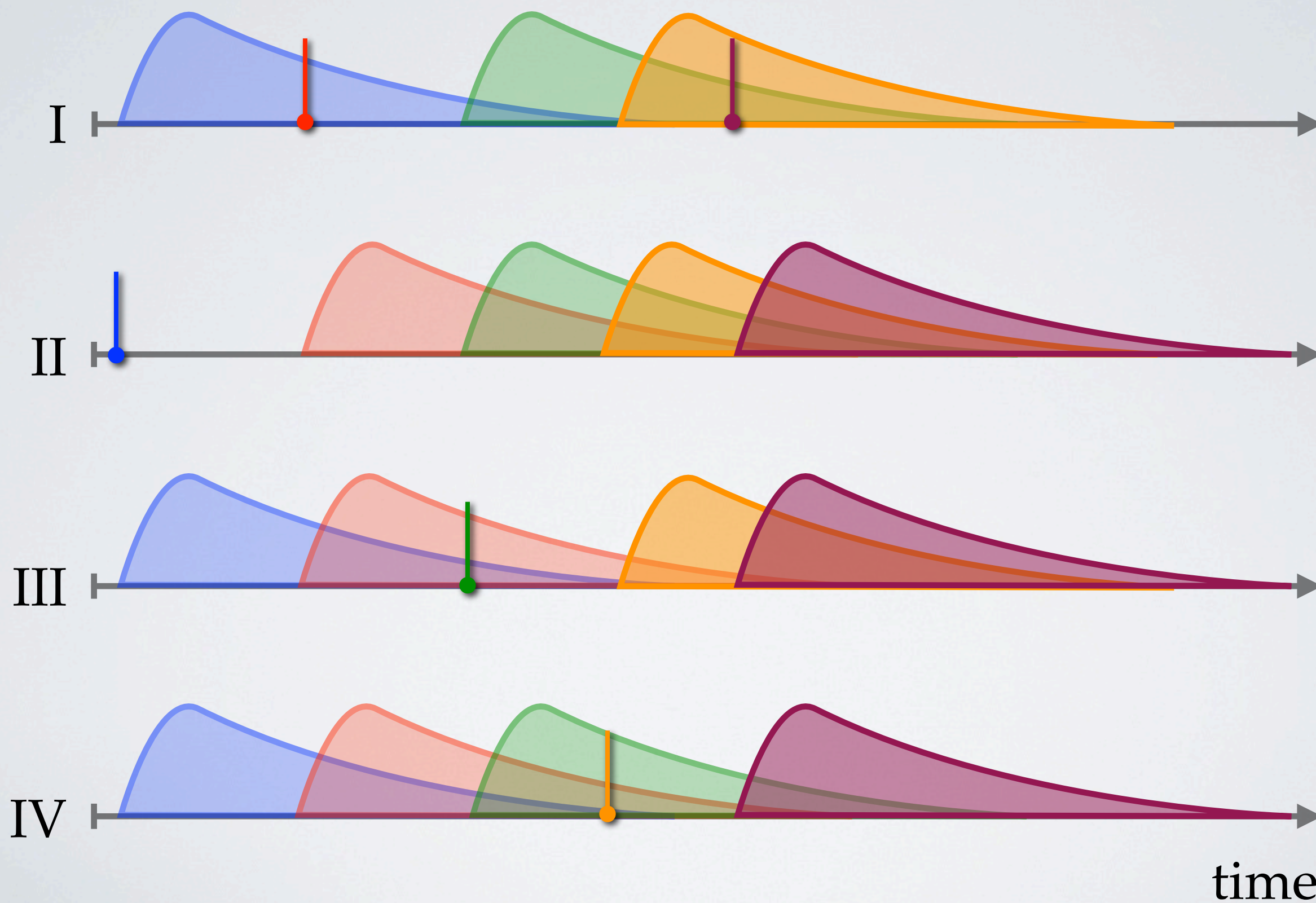


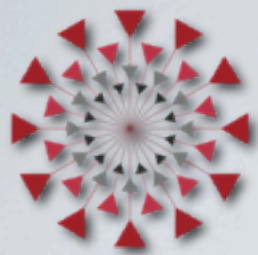
HAWKES PROCESS DYNAMICS



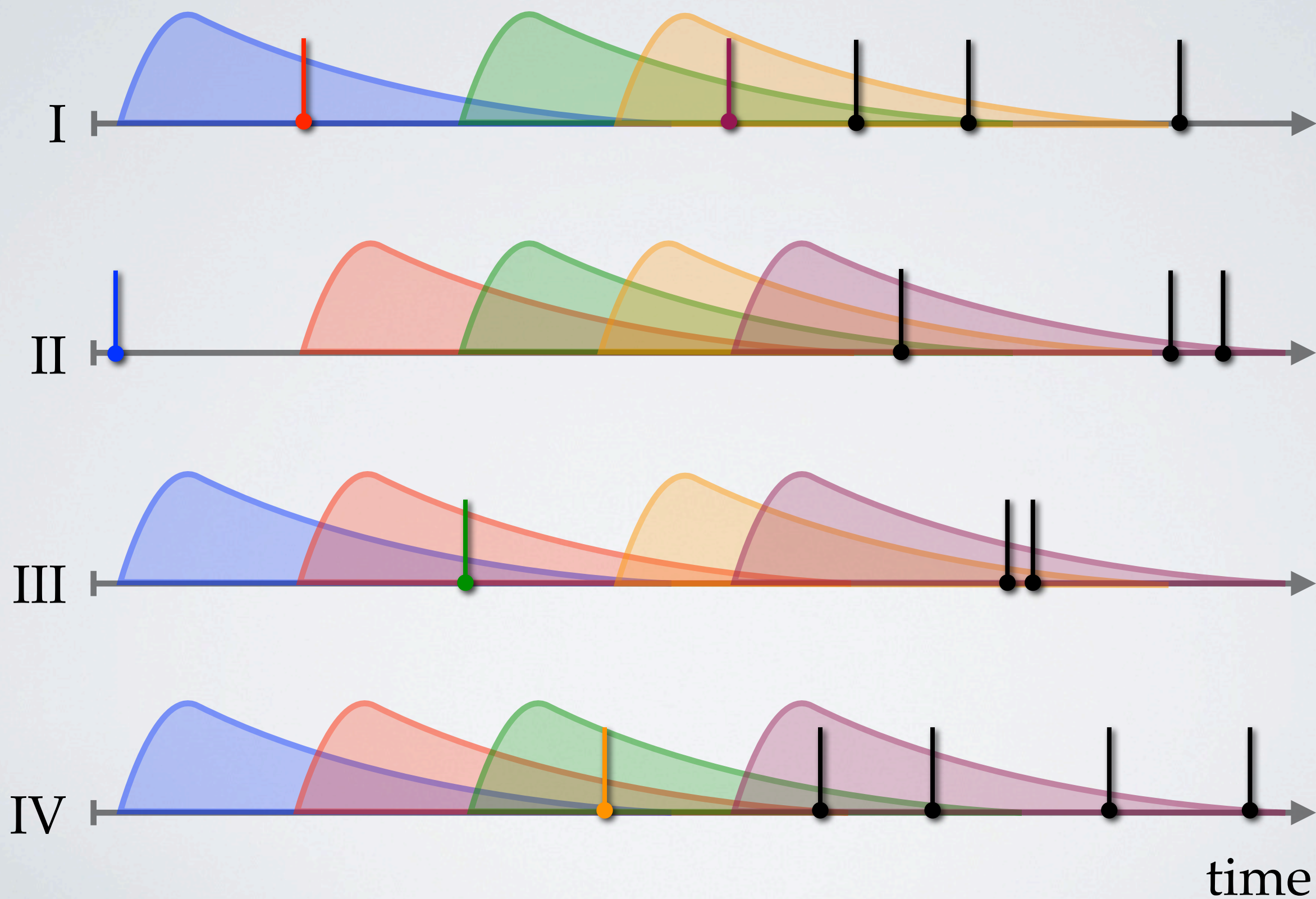


HAWKES PROCESS DYNAMICS



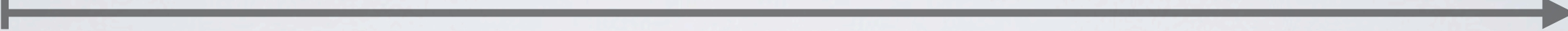


HAWKES PROCESS DYNAMICS

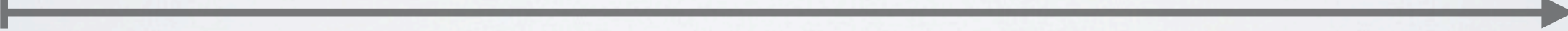


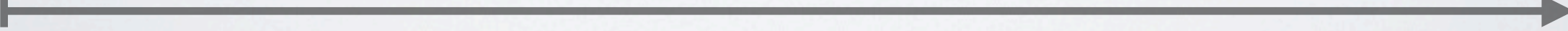


HAWKES PROCESS DYNAMICS

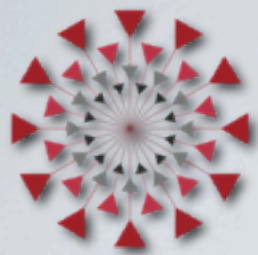
I 

II 

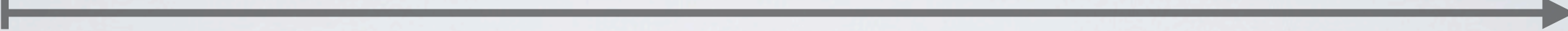
III 

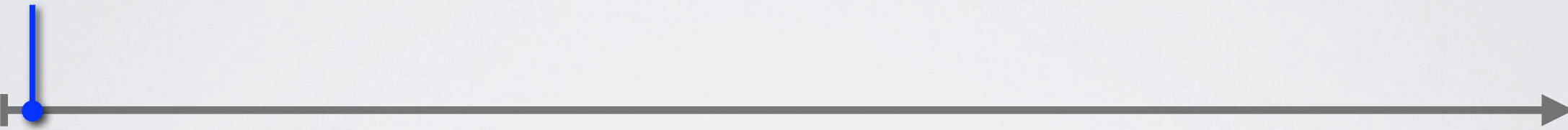
IV 

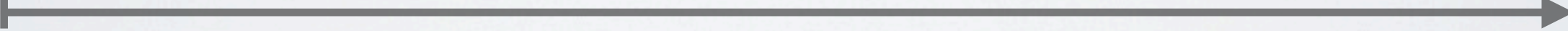
time

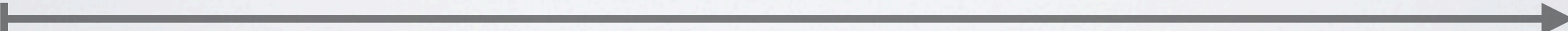


HAWKES PROCESS DYNAMICS

I 

II 

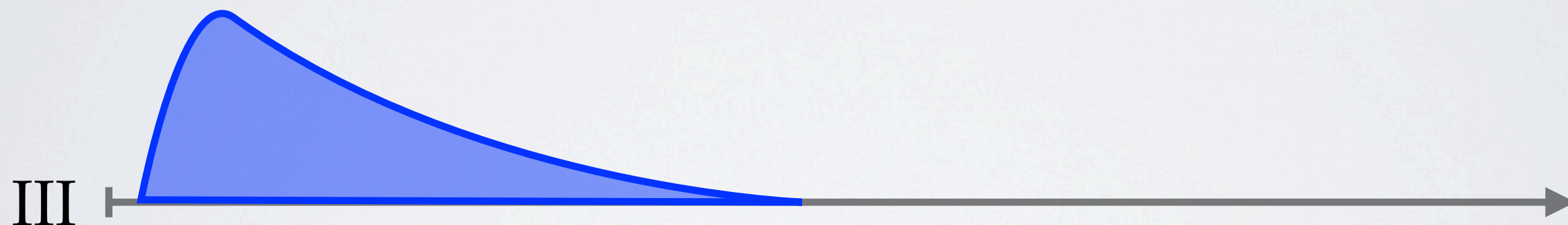
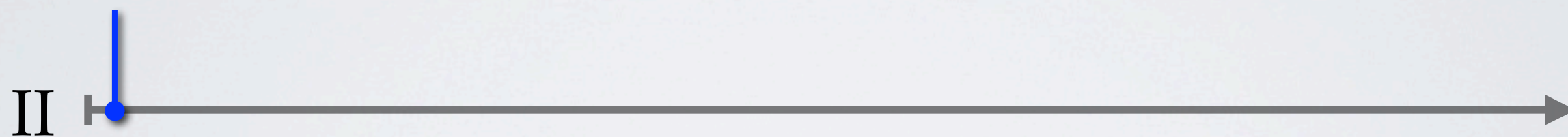
III 

IV 

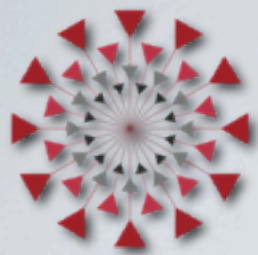
time



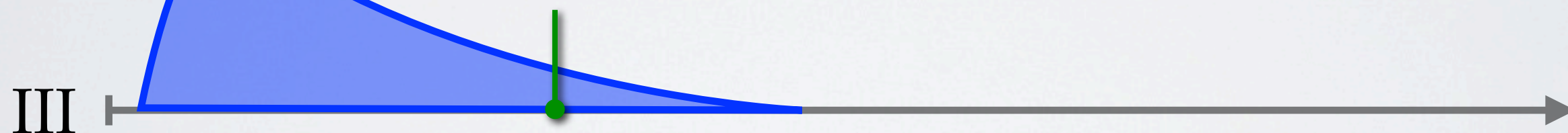
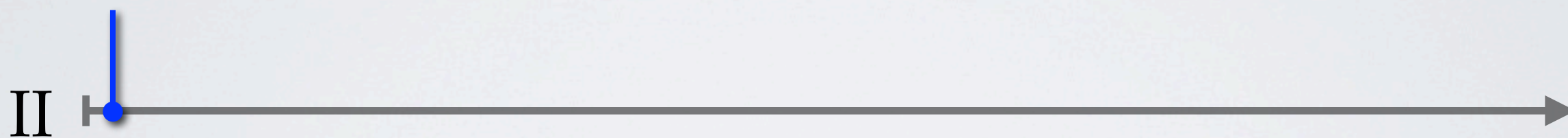
HAWKES PROCESS DYNAMICS



time



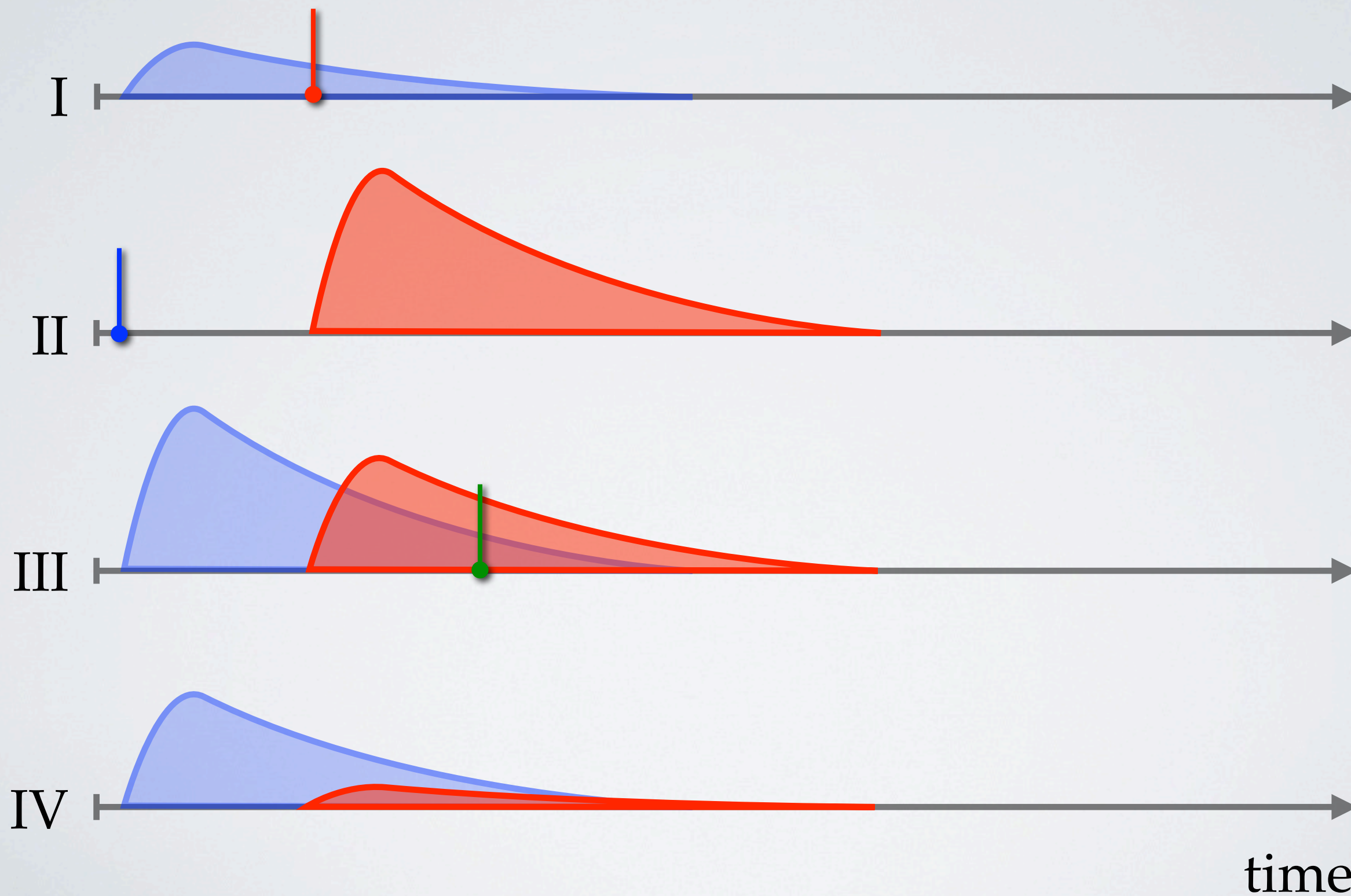
HAWKES PROCESS DYNAMICS

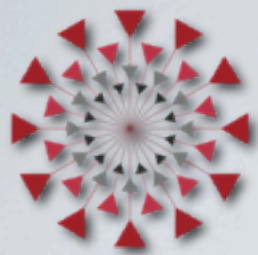


time

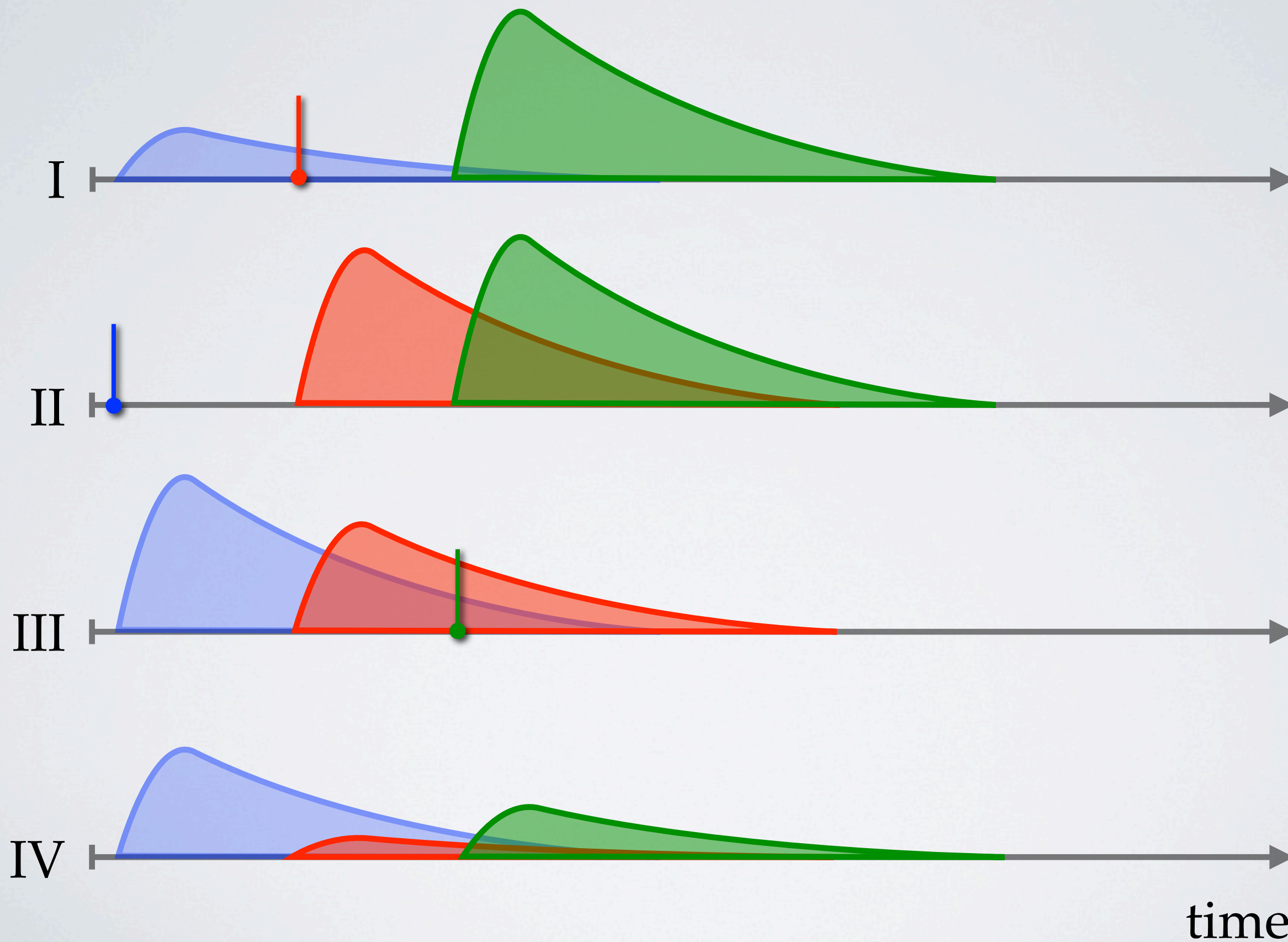


HAWKES PROCESS DYNAMICS



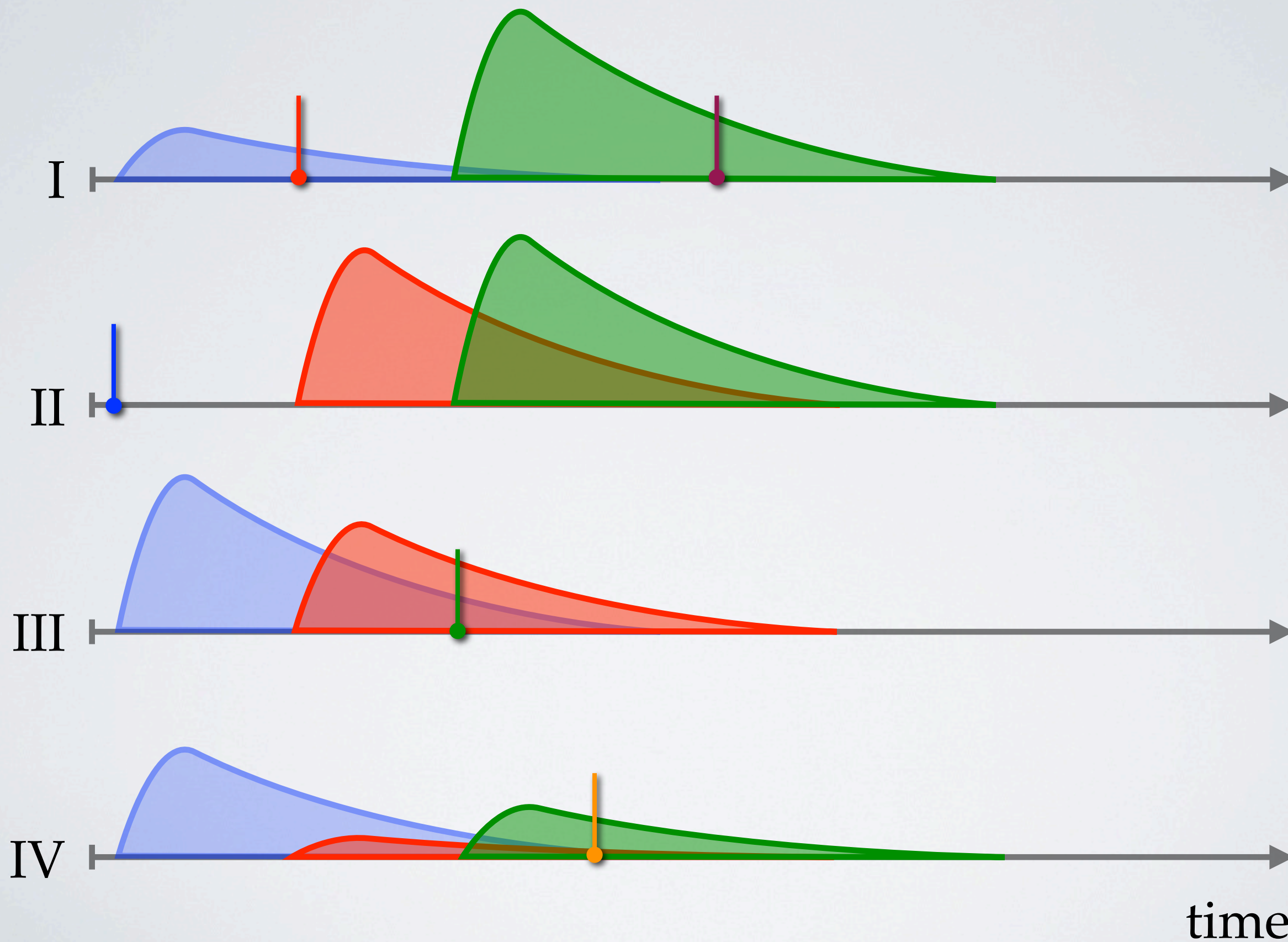


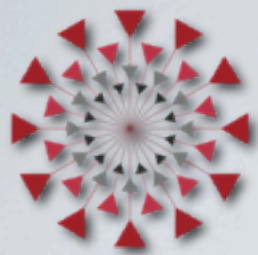
HAWKES PROCESS DYNAMICS



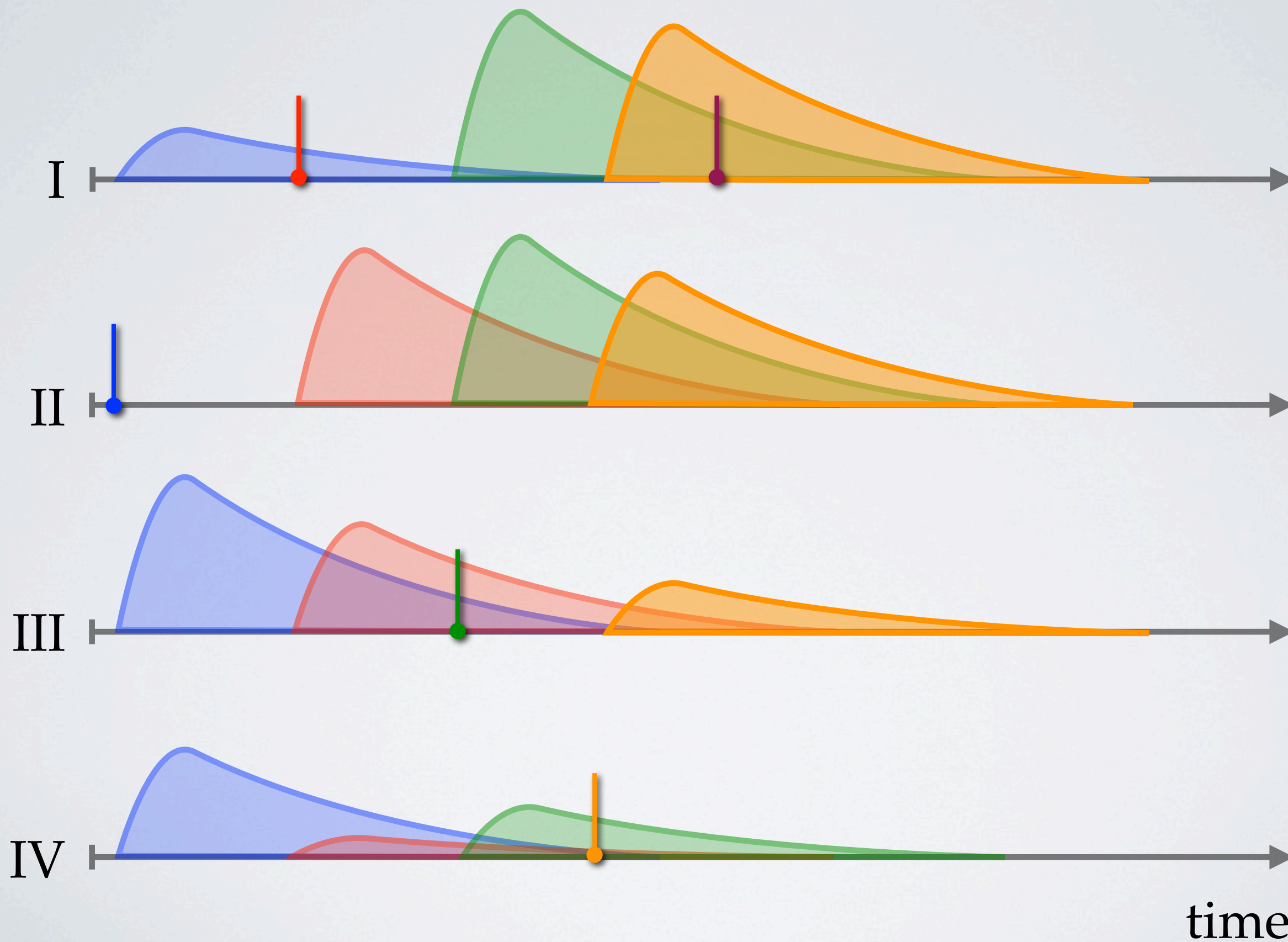


HAWKES PROCESS DYNAMICS



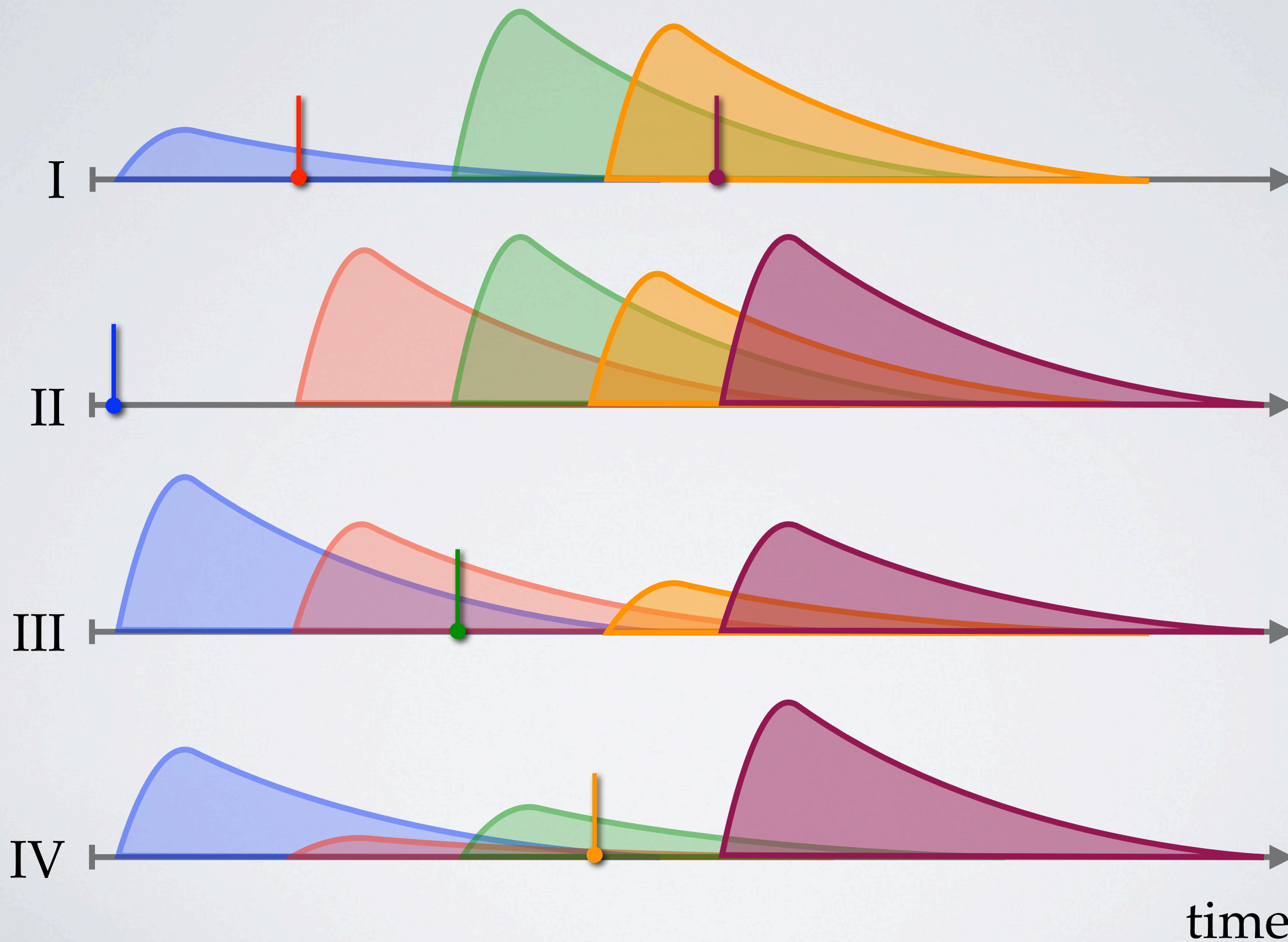


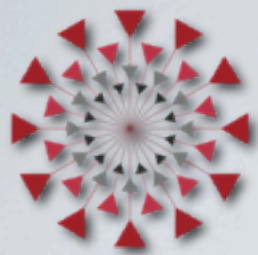
HAWKES PROCESS DYNAMICS



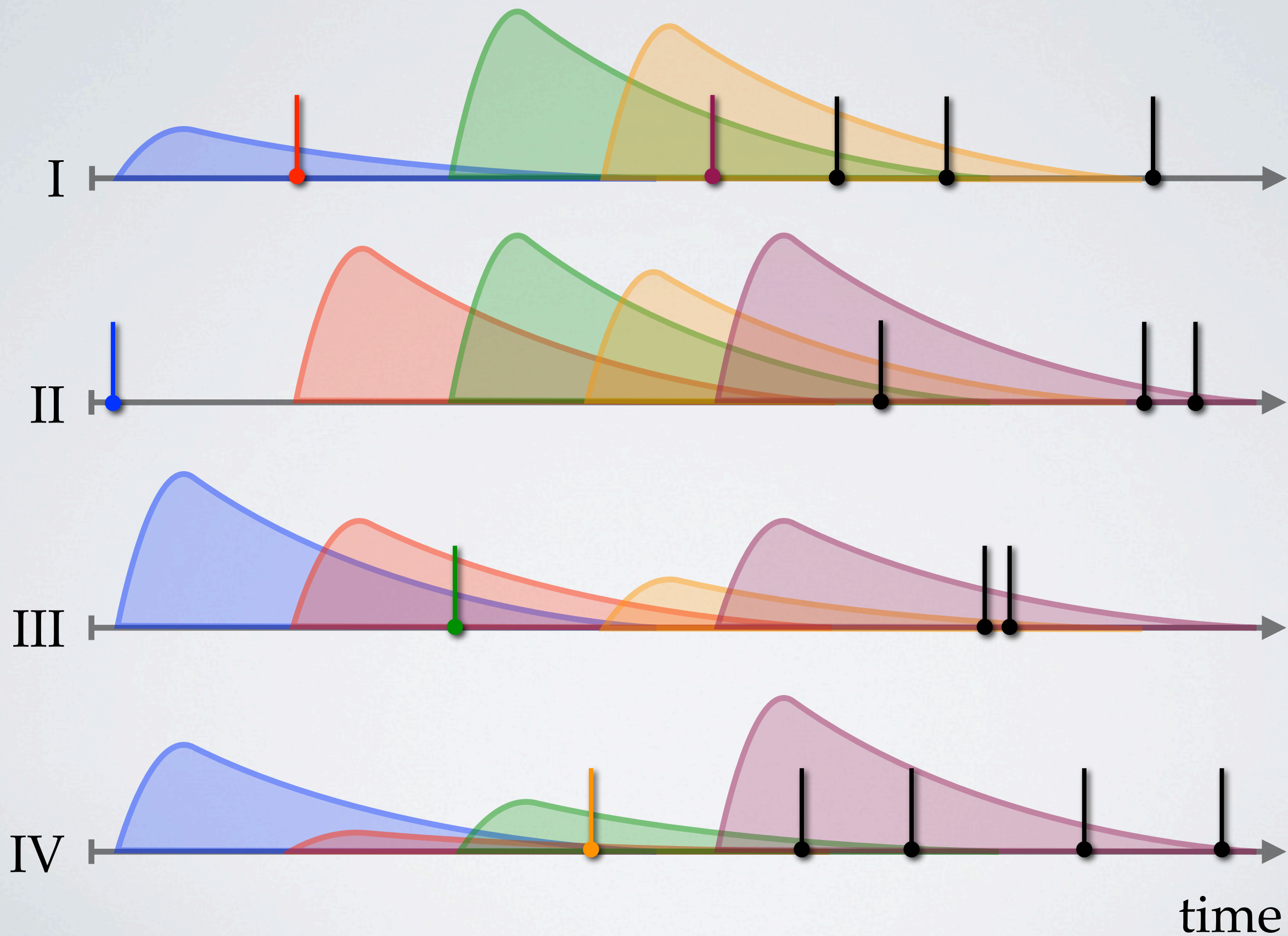


HAWKES PROCESS DYNAMICS





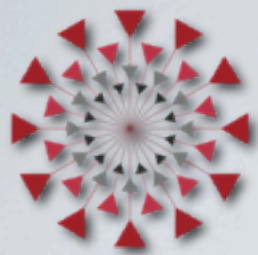
HAWKES PROCESS DYNAMICS



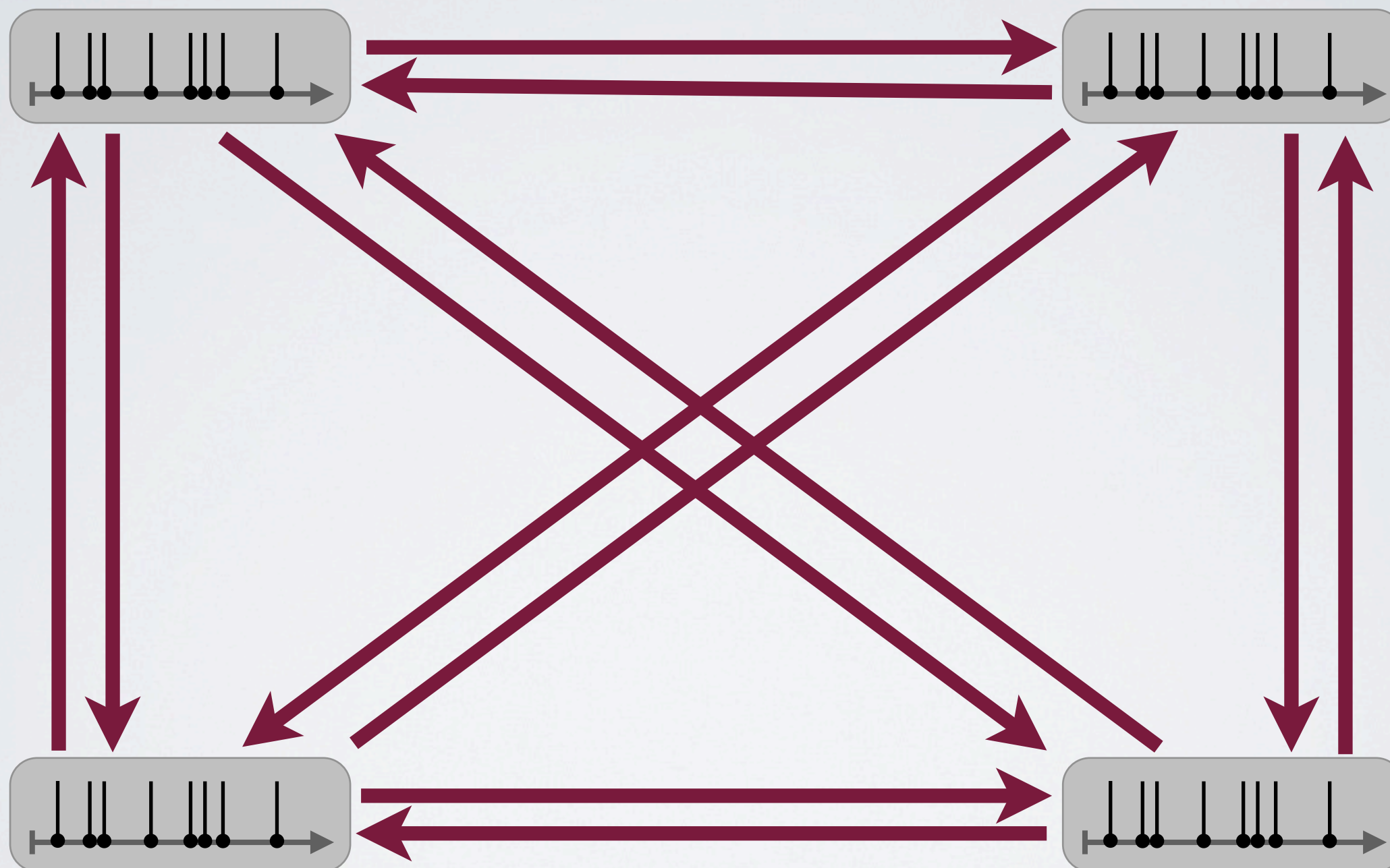


HAWKES PROCESS

- ▶ The Hawkes process specifies conditional Poisson dynamics in causal cascades.
 - ▶ Hawkes - Biometrika (1971), J. RSS-B (1971)
 - ▶ Hawkes - Stochastic Point Processes (1972)
 - ▶ Hawkes & Oakes - J. Applied Prob. (1974)
- ▶ Purely excitatory: each spike increases the intensity according to a weighted temporal kernel.
- ▶ Self-excitation: increase your own rate.
- ▶ Mutual-excitation: increase other rates.
- ▶ Spectral conditions ensure stability.

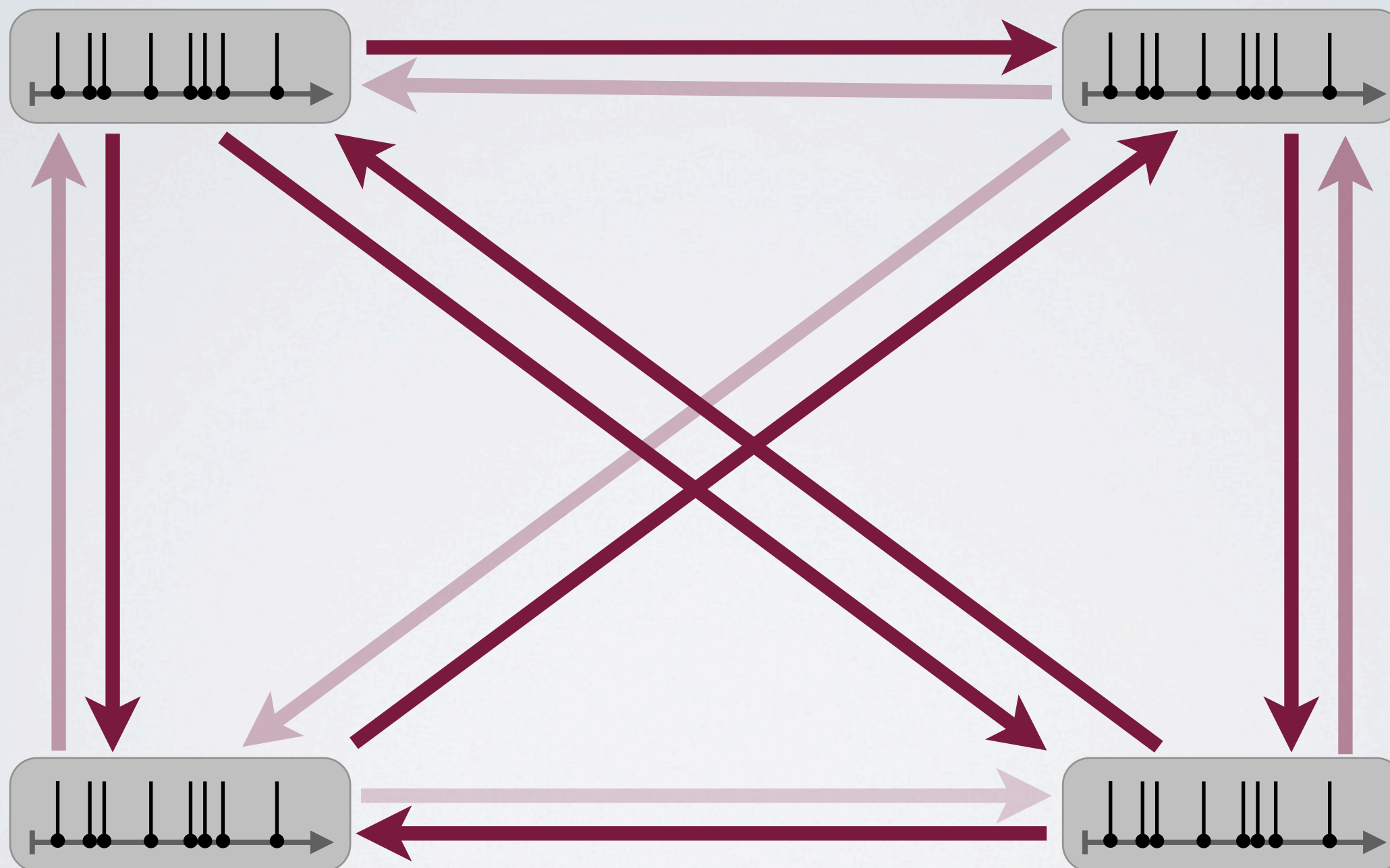


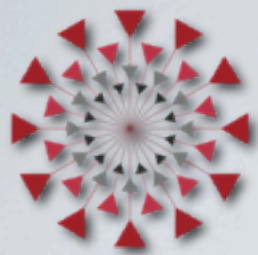
GRAPH STRUCTURE FROM HAWKES DYNAMICS



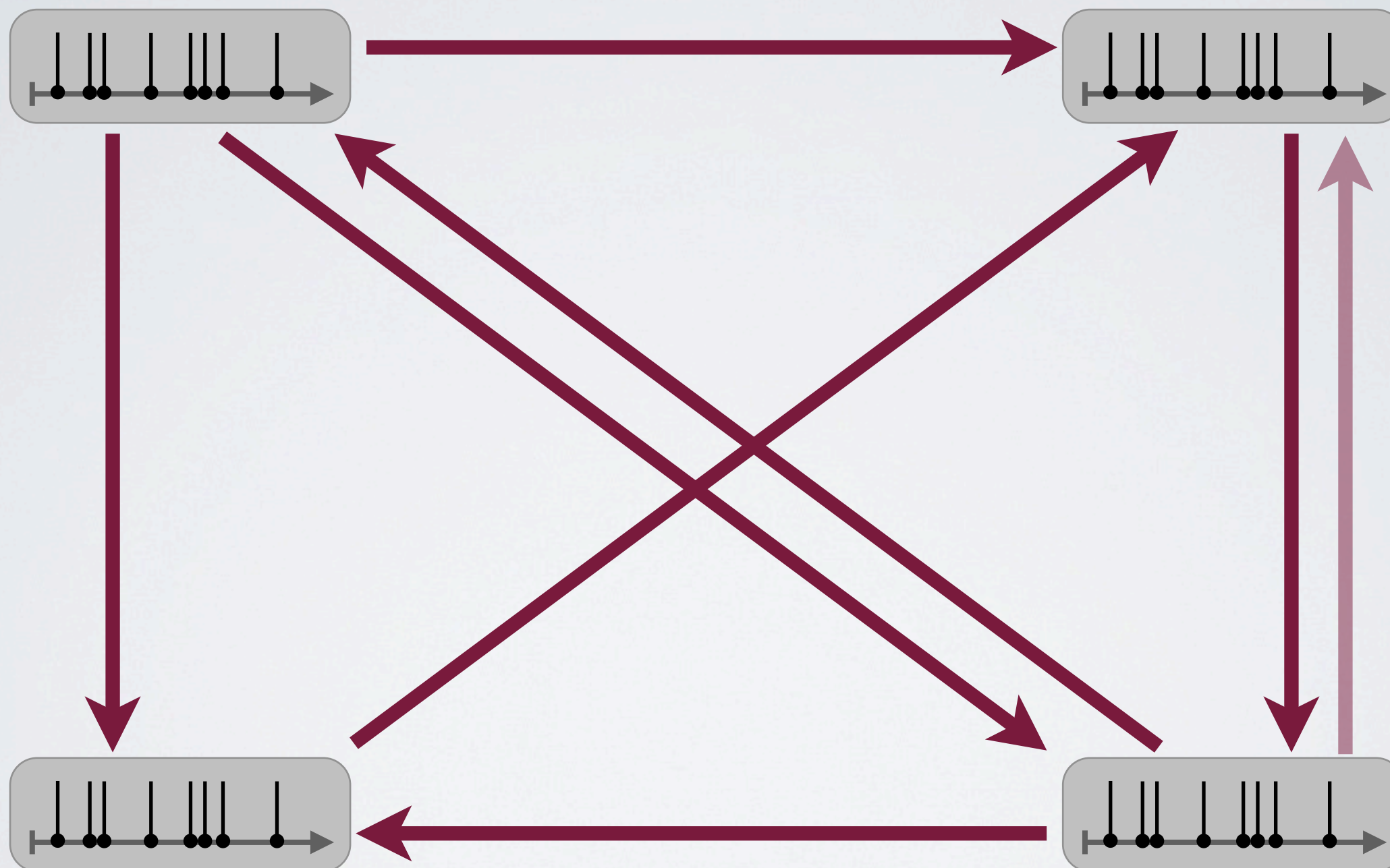


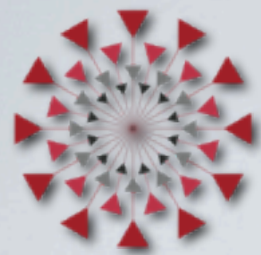
GRAPH STRUCTURE FROM HAWKES DYNAMICS





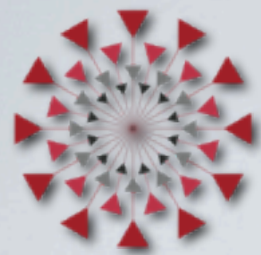
GRAPH STRUCTURE FROM HAWKES DYNAMICS





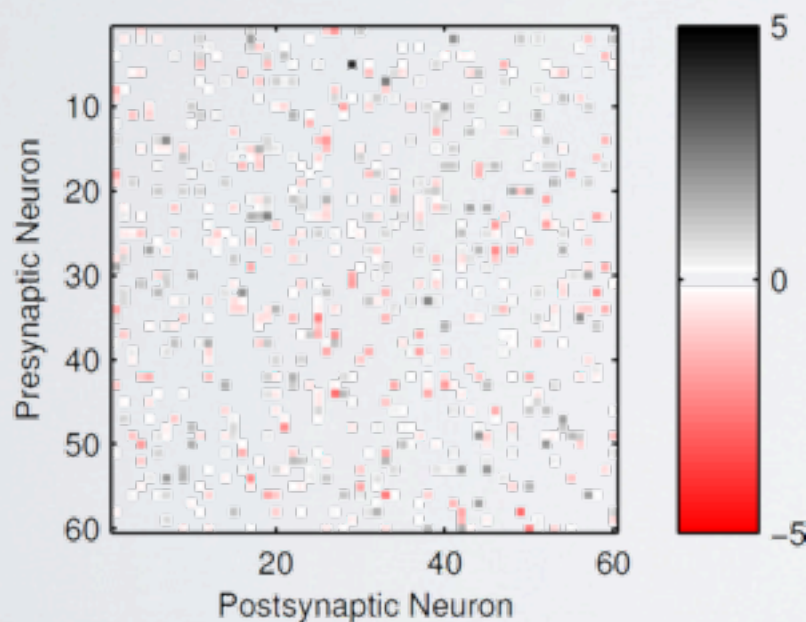
BAYESIAN INFERENCE

- ▶ The Hawkes process provides a likelihood to connect hypotheses about an **unobserved graph** to **observed event data**.
- ▶ With a Bayesian model, we can manipulate the posterior distribution over graphs, marginalizing out nuisance parameters.
- ▶ We can infer the temporal kernels that modulate the interaction.
- ▶ We can perform model comparison between different random graph models and their properties.

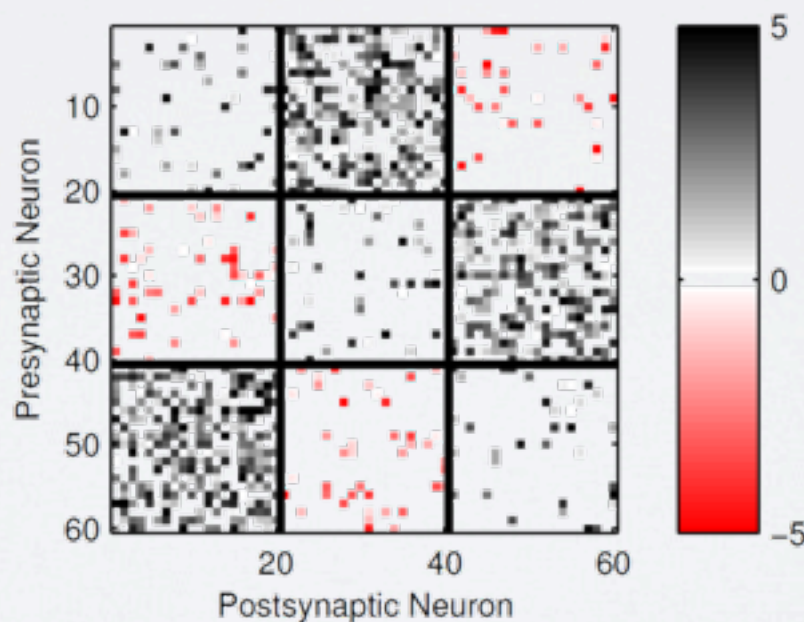


EXCHANGEABLE RANDOM GRAPHS

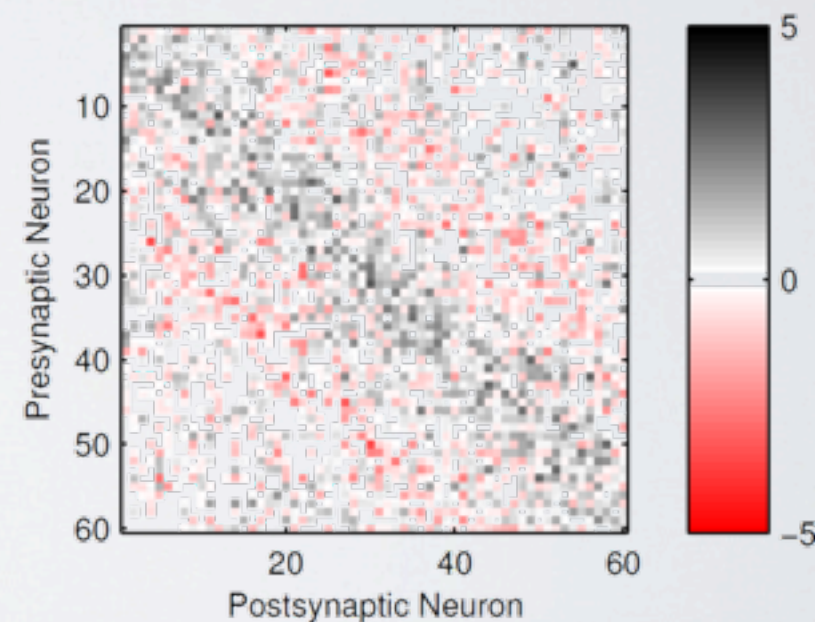
- ▶ Recent work on exchangeable random graphs provides a rich set of priors for underlying networks, e.g., Diaconis & Janson (2007), Orbanz & Roy (2013).
- ▶ Aldous-Hoover unifies many existing graph models:



Erdős-Renyi



Stochastic Block
Model

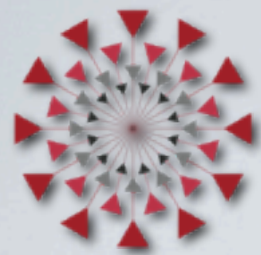


Latent Distance
Model



HAWKES PROCESS FORMALISM

- ▶ K nodes with events in $[0, T]$.
- ▶ N ordered event times $s_n \in [0, T]$.
- ▶ Node of event n given by $c_n \in \{1, 2, \dots, K\}$.
- ▶ Base rates $\lambda_k^0(t)$
- ▶ Kernel $g_\theta(t)$, such that $g_\theta(t < 0) = 0$ and $\int_0^\infty g_\theta(t) dt = 1$.
- ▶ Binary adjacency matrix $\mathbf{A} \in \{0, 1\}^{K \times K}$
- ▶ Interaction weight matrix $\mathbf{W} \in \mathbb{R}_+^{K \times K}$
- ▶ An event on k induces $W_{k,k'} A_{k,k'}$ expected events on k' .



HAWKES PROCESS WITH DATA AUGMENTATION

- ▶ Instantaneous Poisson rate:

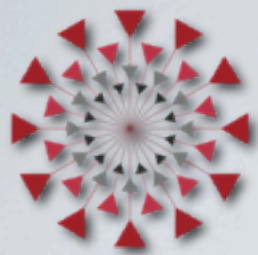
$$\lambda_k(t) = \lambda_k^0(t) + \sum_{n=1}^N \mathbb{I}(s_n < t) A_{c_n, k} W_{c_n, k} g_\theta(t - s_n)$$

- ▶ The superposition property of Poisson processes means that each event is explained by either the background rate or exactly one previous event.
- ▶ Use $\mathbf{Z} \in \{0, 1\}^{N \times N}$ to represent these latent variables, where $Z_{n, n'} = 1$ means that event n induced event n' .



DATA-AUGMENTED LIKELIHOOD

$$\begin{aligned} p(\{s_n, c_n\}_{n=1}^N, \mathbf{Z} \mid \mathbf{A}, \mathbf{W}, \{\lambda_k^0(t)\}_{k=1}^K, \theta) = \\ \prod_{k=1}^K \exp \left\{ - \int_0^T \lambda_k^0(\tau) d\tau \right\} \prod_{n=1}^N \lambda_k^0(s_n)^{\mathbb{I}(c_n=k) \mathbb{I}(1 - \sum_{n'} Z_{n',n})} \\ \times \prod_{k'=1}^K \left\{ - \int_{s_n}^T A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n) d\tau \right\} \\ \times \prod_{n'=1}^N (A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n))^{Z_{n,n'}} \end{aligned}$$



DATA-AUGMENTED LIKELIHOOD

$$p(\{s_n, c_n\}_{n=1}^N, \mathbf{Z} \mid \mathbf{A}, \mathbf{W}, \{\lambda_k^0(t)\}_{k=1}^K, \theta) =$$
$$\prod_{k=1}^K \exp \left\{ - \int_0^T \lambda_k^0(\tau) d\tau \right\} \prod_{n=1}^N \lambda_k^0(s_n)^{\mathbb{I}(c_n=k) \mathbb{I}(1 - \sum_{n'} Z_{n',n})}$$
$$\times \prod_{k'=1}^K \left\{ - \int_{s_n}^T A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n) d\tau \right\}$$
$$\times \prod_{n'=1}^N (A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n))^{Z_{n,n'}}$$

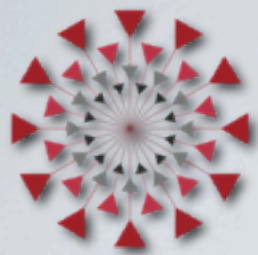
Poisson process likelihood for events
from the background process.



DATA-AUGMENTED LIKELIHOOD

$$\begin{aligned} p(\{s_n, c_n\}_{n=1}^N, \mathbf{Z} \mid \mathbf{A}, \mathbf{W}, \{\lambda_k^0(t)\}_{k=1}^K, \theta) = \\ \prod_{k=1}^K \exp \left\{ - \int_0^T \lambda_k^0(\tau) d\tau \right\} \prod_{n=1}^N \lambda_k^0(s_n)^{\mathbb{I}(c_n=k) \mathbb{I}(1 - \sum_{n'} Z_{n',n})} \\ \times \prod_{k'=1}^K \left\{ - \int_{s_n}^T A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n) d\tau \right\} \\ \times \prod_{n'=1}^N (A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n))^{Z_{n,n'}} \end{aligned}$$

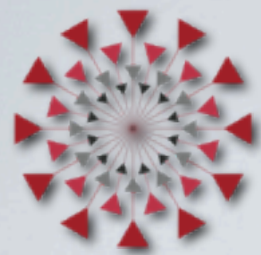
Poisson process likelihood for events induced by previous events.



DATA-AUGMENTED LIKELIHOOD

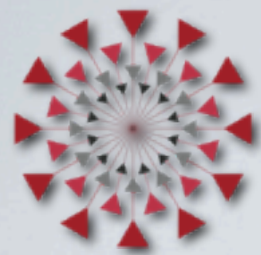
$$\begin{aligned} p(\{s_n, c_n\}_{n=1}^N, \mathbf{Z} \mid \mathbf{A}, \mathbf{W}, \{\lambda_k^0(t)\}_{k=1}^K, \theta) = \\ \prod_{k=1}^K \exp \left\{ - \int_0^T \lambda_k^0(\tau) d\tau \right\} \prod_{n=1}^N \lambda_k^0(s_n)^{\mathbb{I}(c_n=k) \mathbb{I}(1 - \sum_{n'} Z_{n',n})} \\ \times \prod_{k'=1}^K \left\{ - \int_{s_n}^T A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n) d\tau \right\} \\ \times \prod_{n'=1}^N (A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n))^{Z_{n,n'}} \end{aligned}$$

Ugly looking, but just normalization constants.



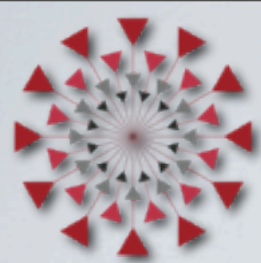
MODELING DETAILS

- ▶ Logistic-normal impulse response
 - ▶ *Reasonably flexible, with compact support.*
- ▶ Gaussian process for log background rate
 - ▶ *Smoothly-varying or periodic external effects.*
- ▶ Conjugate gamma priors for weights
 - ▶ *Can also be coupled via, e.g., latent block identity.*



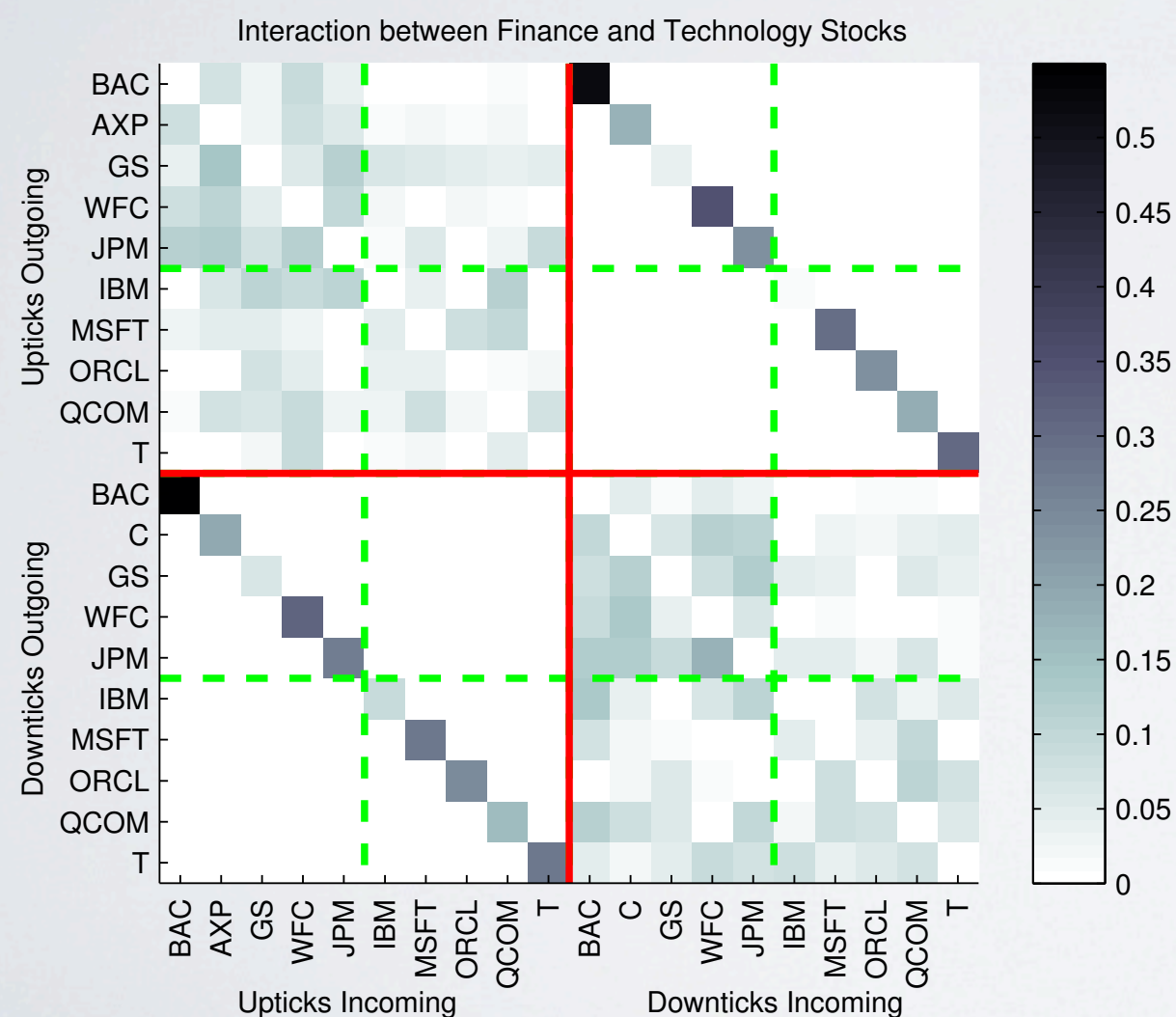
INFERENCE WITH MCMC

- ▶ Graph structure: collapsed block Gibbs
 - ▶ Edge weights: Gibbs (conjugate gamma posterior)
 - ▶ Latent parent explanations: parallel Gibbs
 - ▶ Background rates: elliptical slice sampling
 - ▶ Temporal kernels: Gibbs or slice sampling
-
- ▶ Many transitions can be efficiently simulated on a GPU, enabling models with hundreds of nodes and millions of events.

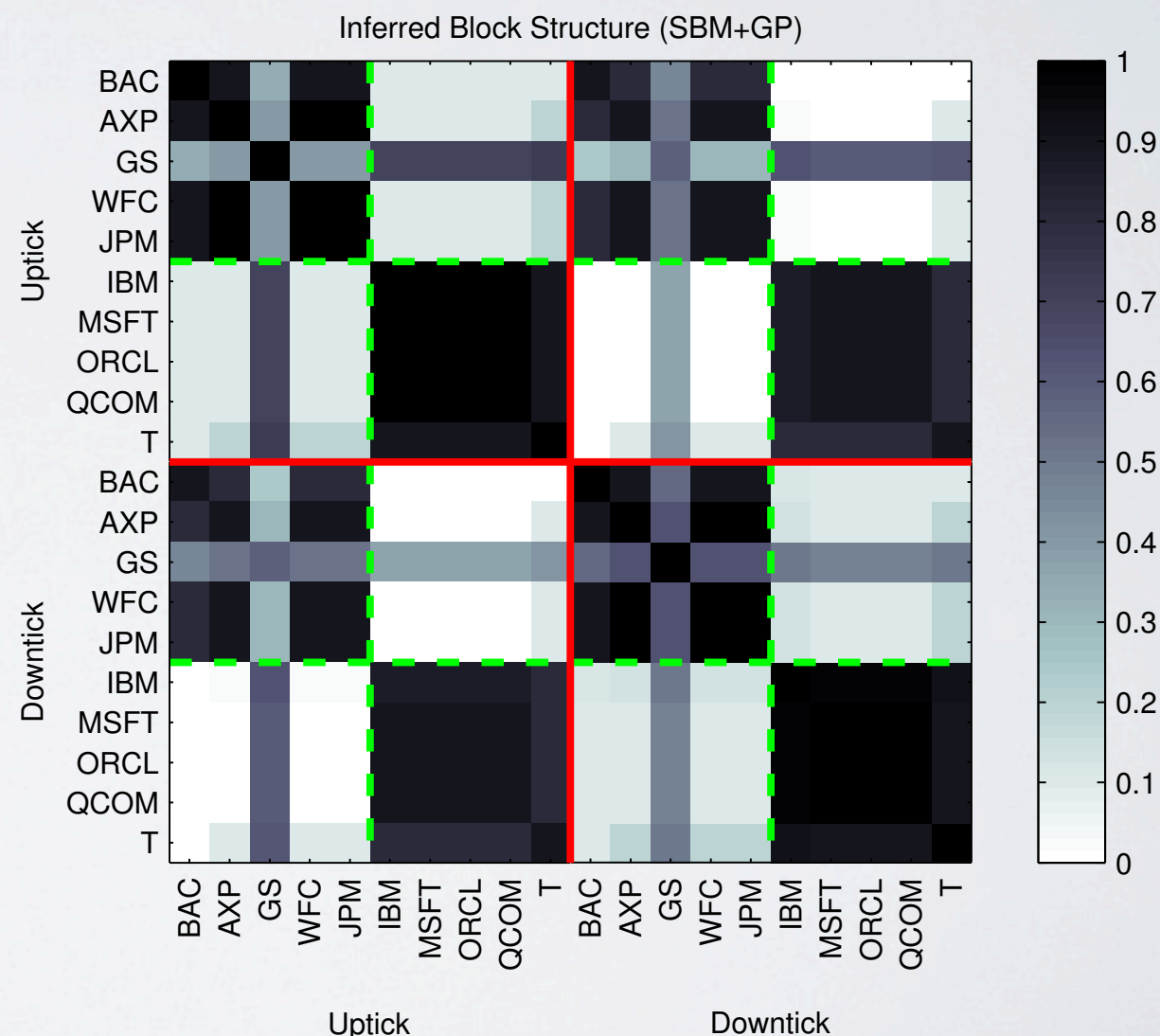


FINANCIAL UPTICKS/DOWNTICKS

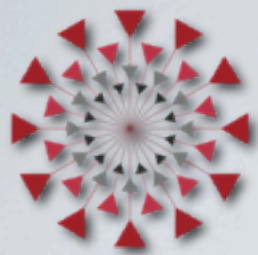
10 stocks, finance and tech sectors, intraday price moves over one week in Sept 2009.



Weighted Edges

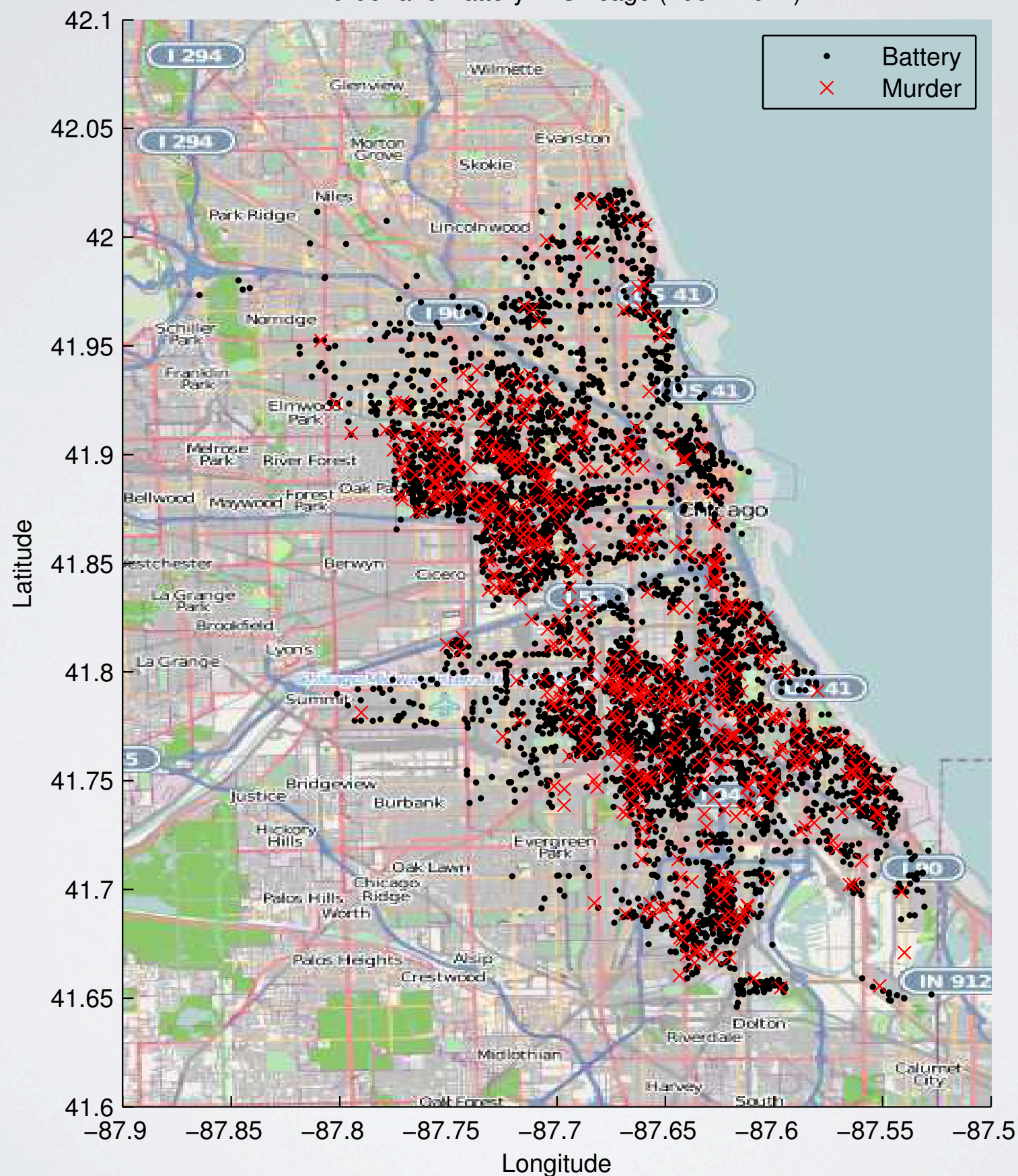


Block Membership



HOMICIDES AND ARMED BATTERIES IN CHICAGO

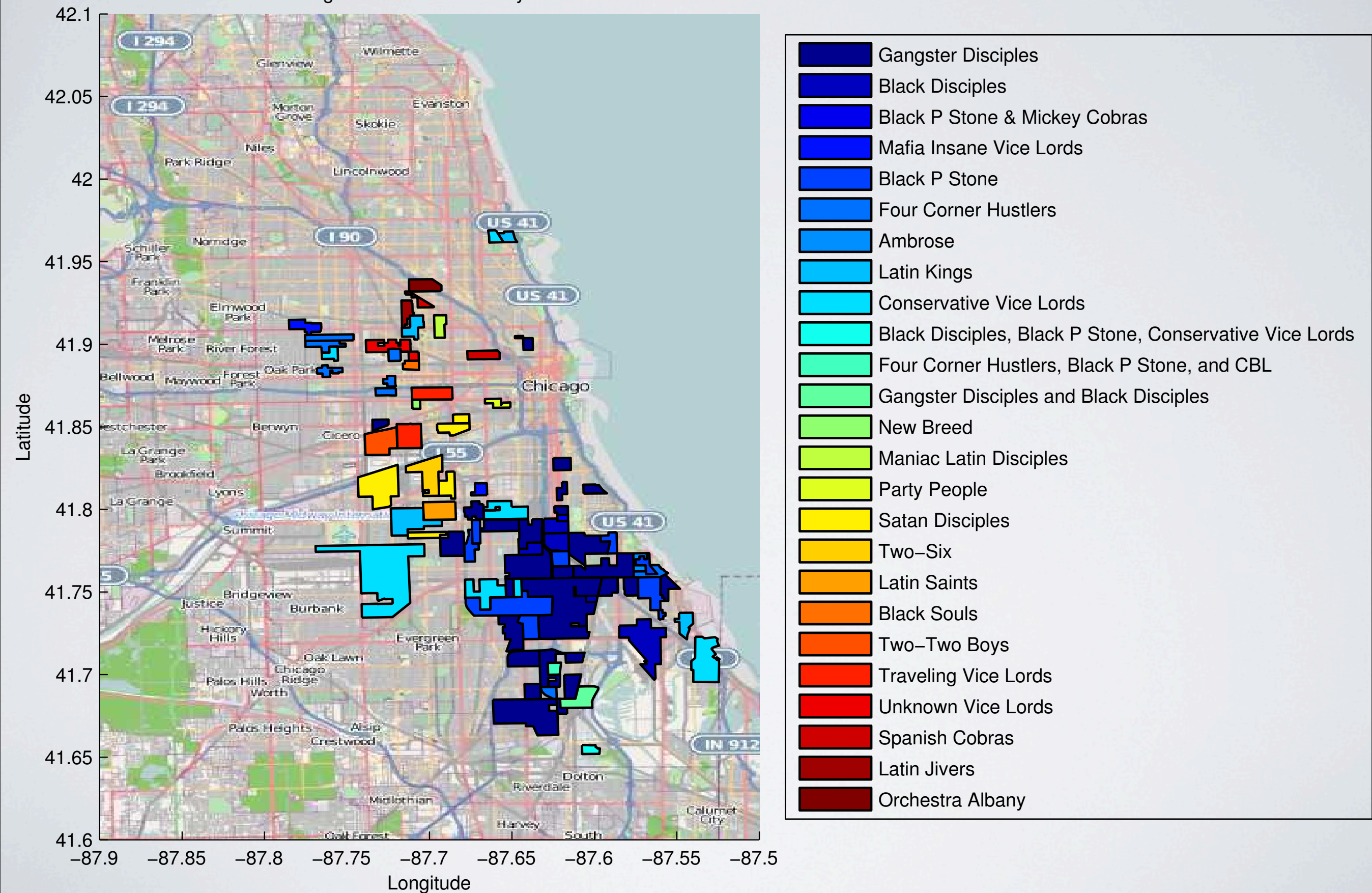
Murder and Battery in Chicago (2001–2012)

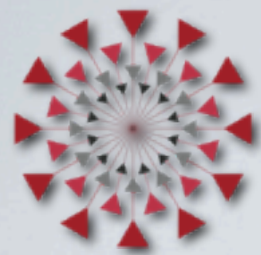




CPD-IDENTIFIED GANNG BOUNDARIES

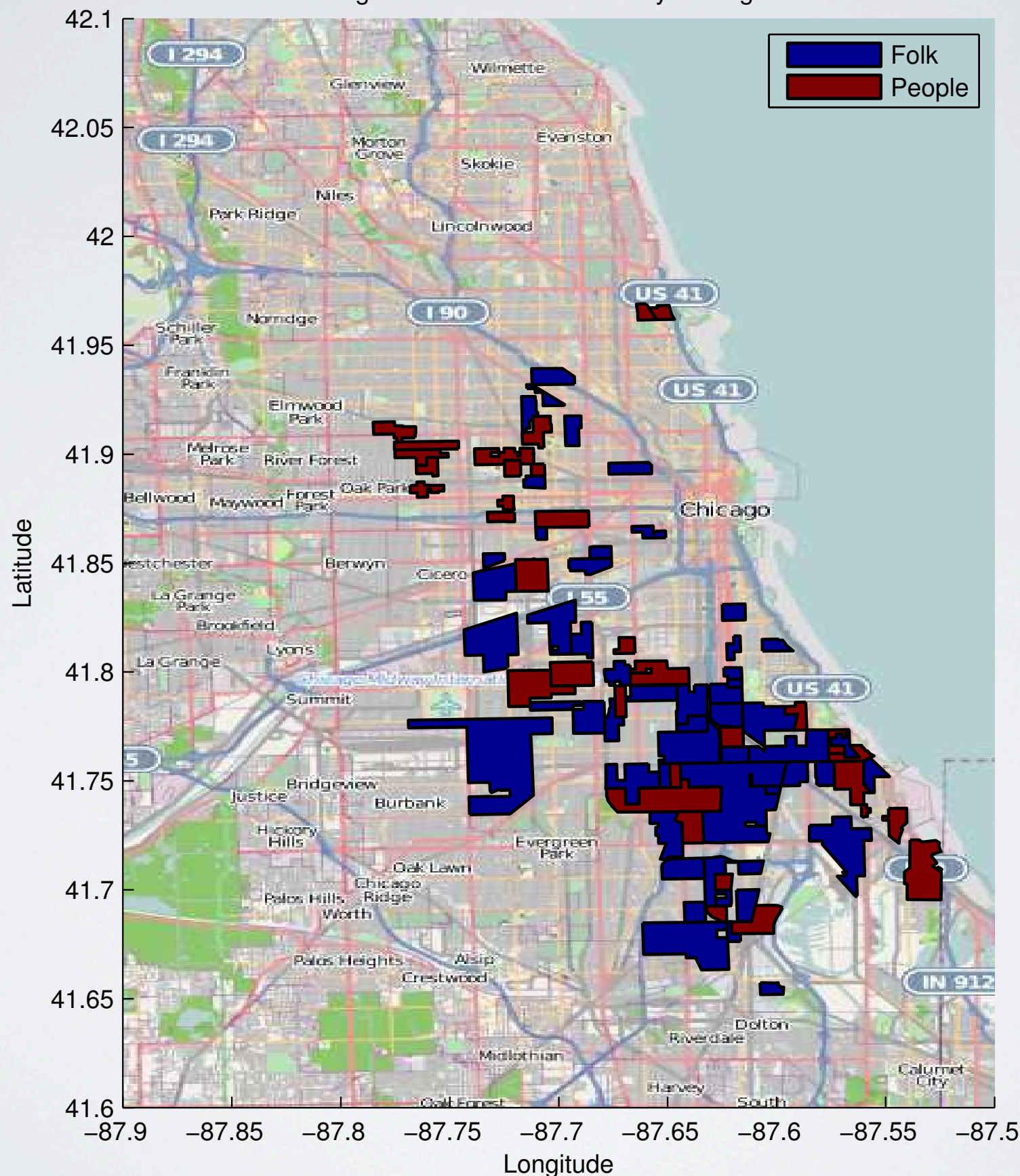
Gang territories identified by CPD





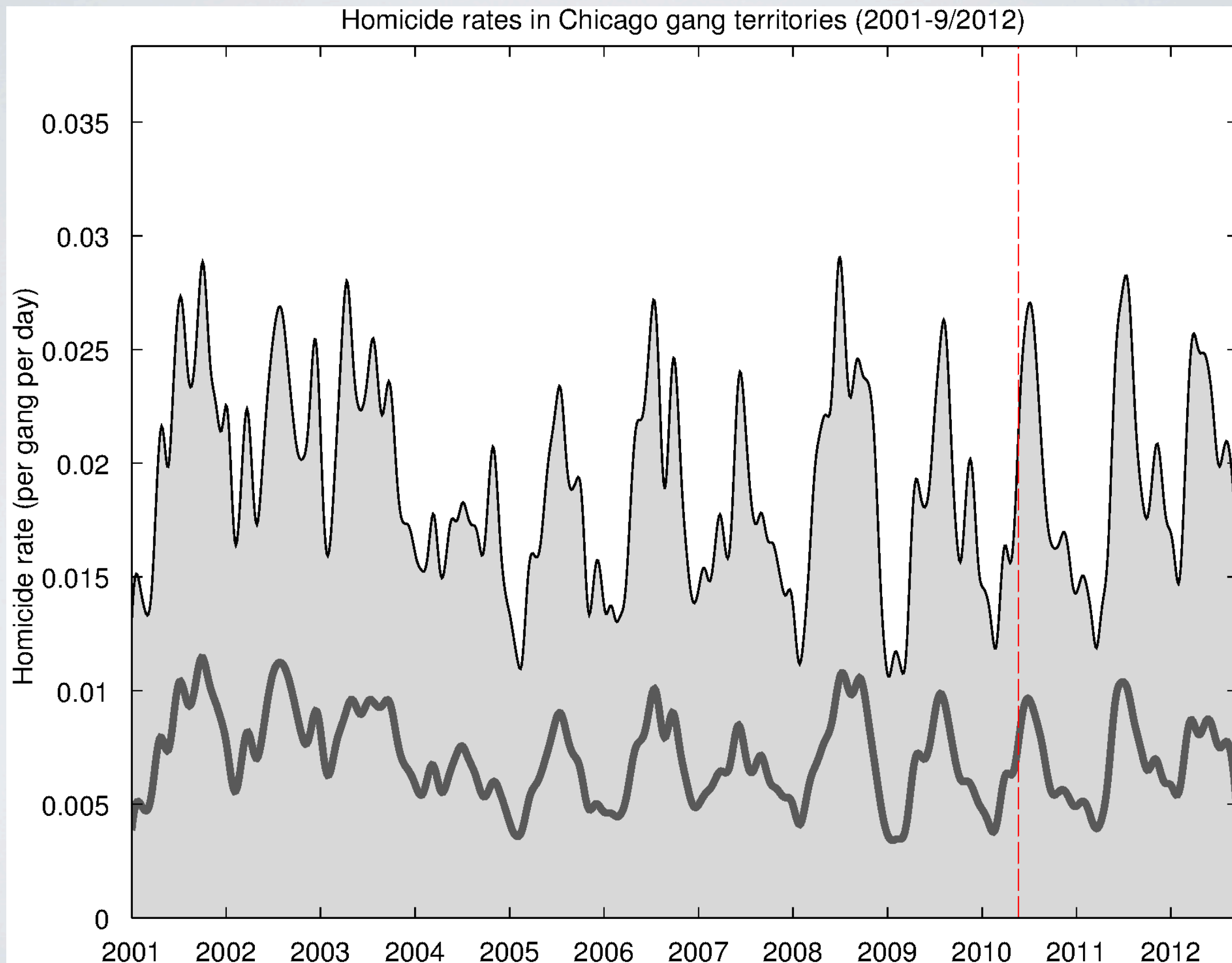
ALLIANCES: FOLK NATION VS. PEOPLE NATION

Gang Affiliations as Identified by Chicago PD





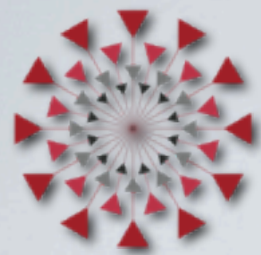
SEASONAL VARIATION





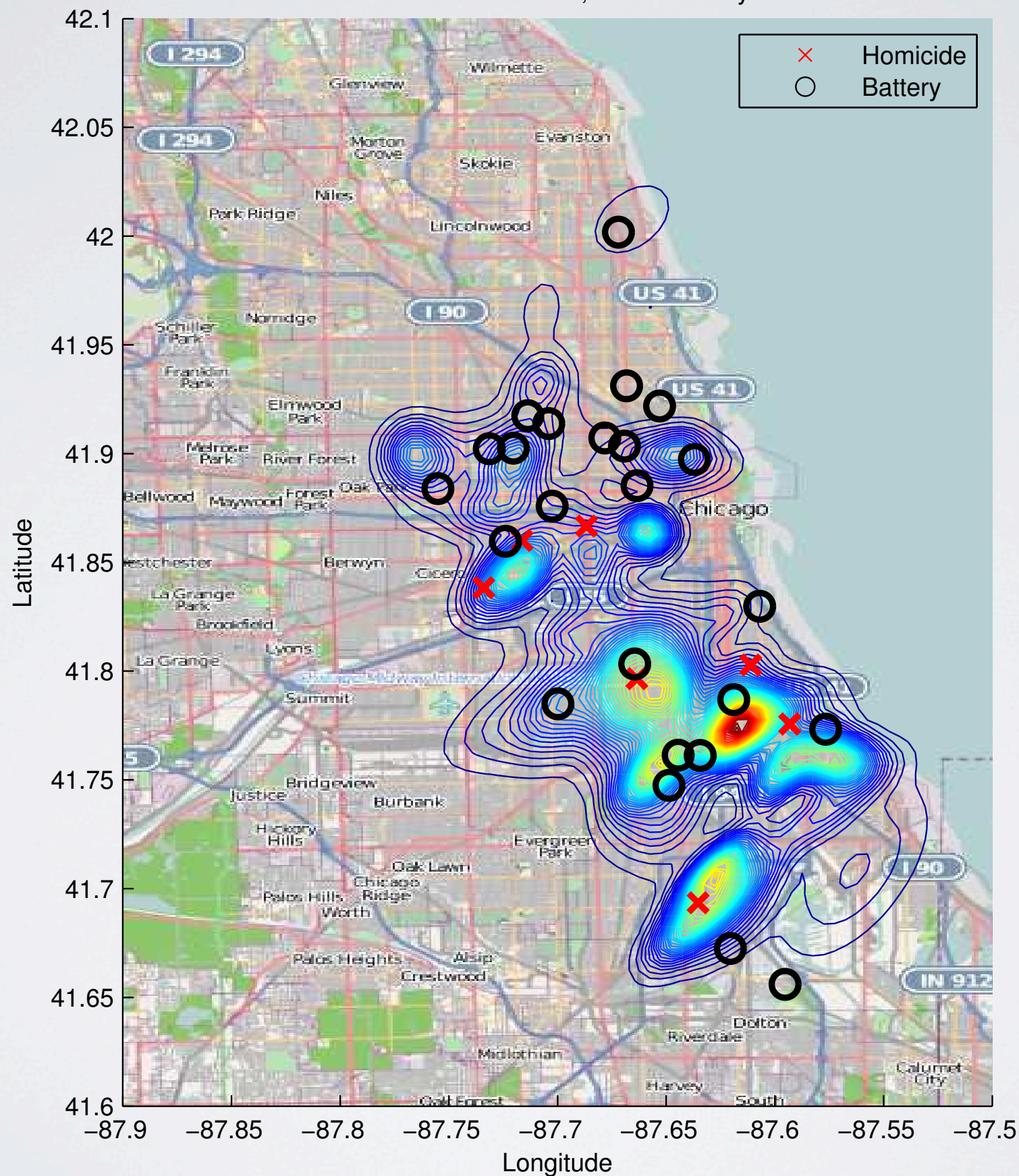
We also in sources, u model an proce

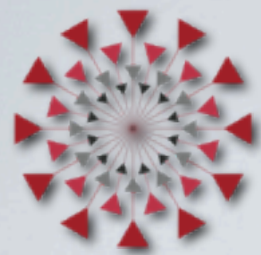
We also infer the event sources, using a spatial model and a Dirichlet process prior.



TRYING TO PREDICT HOTSPOTS

Predicted Homicide Rate, Memorial Day 2012



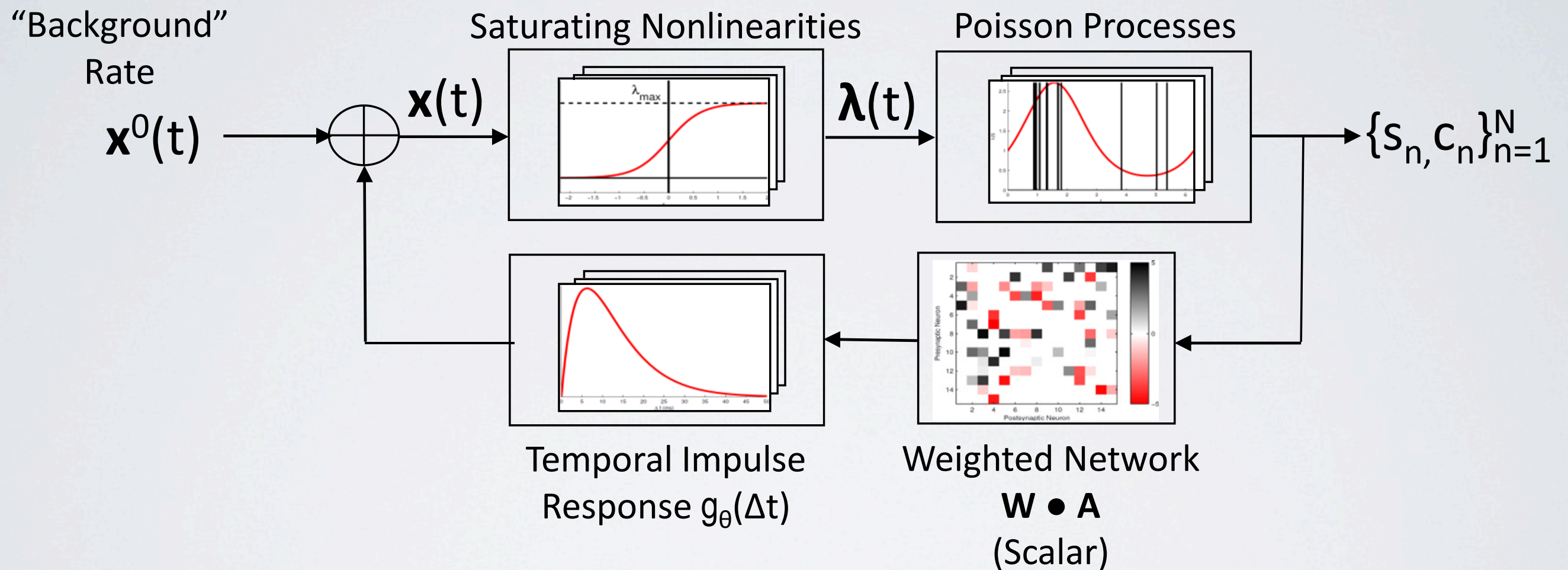


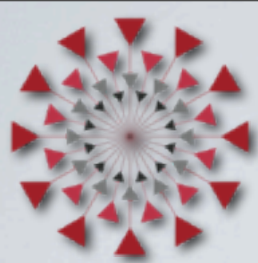
NEURAL MODELING WITH INHIBITION

- ▶ Neural functional connectivity involves both **excitation** and **inhibition**, so the pure Hawkes is inappropriate.
- ▶ We extend the model to allow for negative weights and include a saturating nonlinearity.
- ▶ This effective becomes a Bayesian variant of the popular generalized linear model (GLM) from the computational neuroscience literature.
- ▶ We can leverage graph priors within this framework to discover latent neural properties and connectivity.



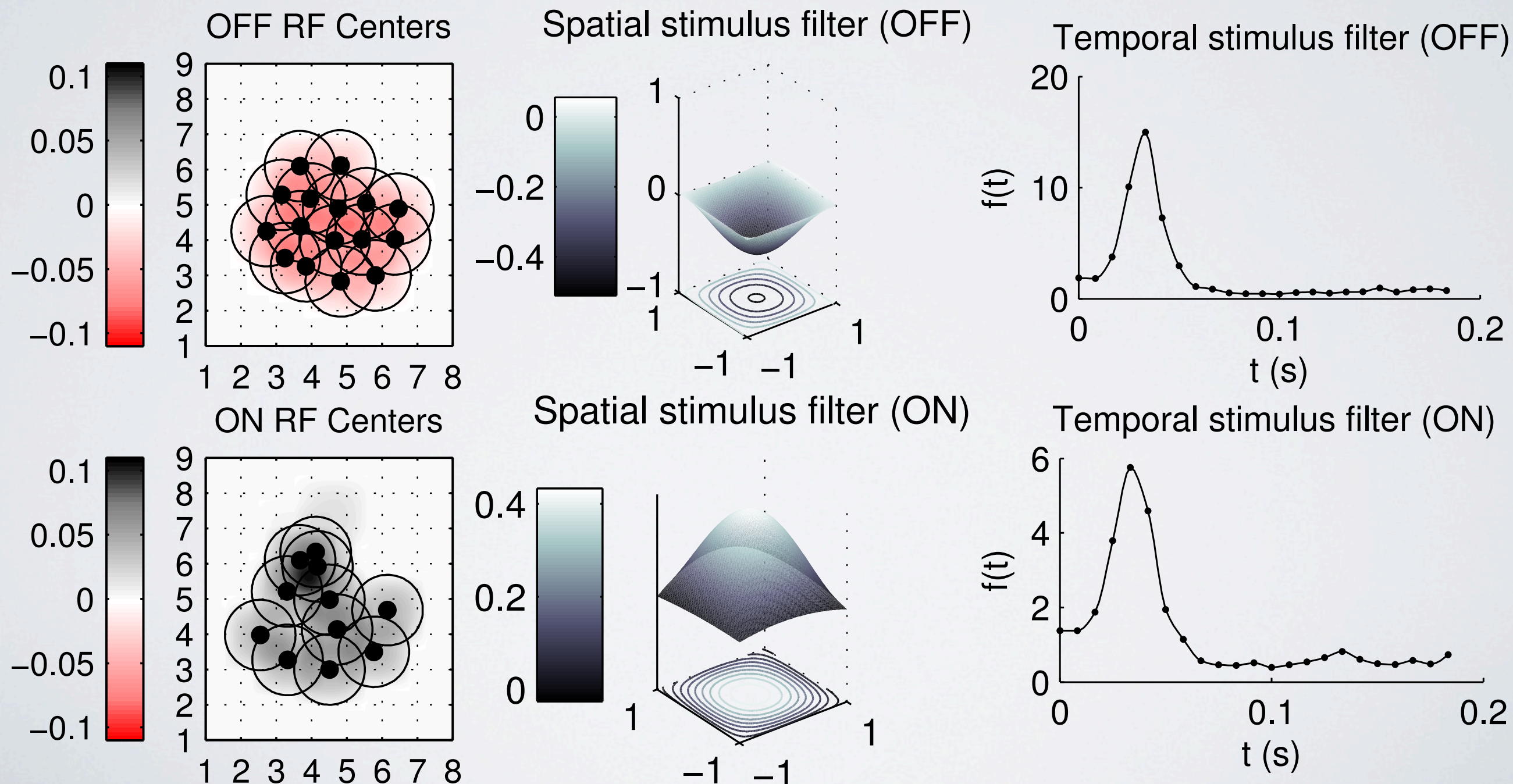
BAYESIAN GLM FOR NEURAL DATA





RETINAL GANGLION CELLS

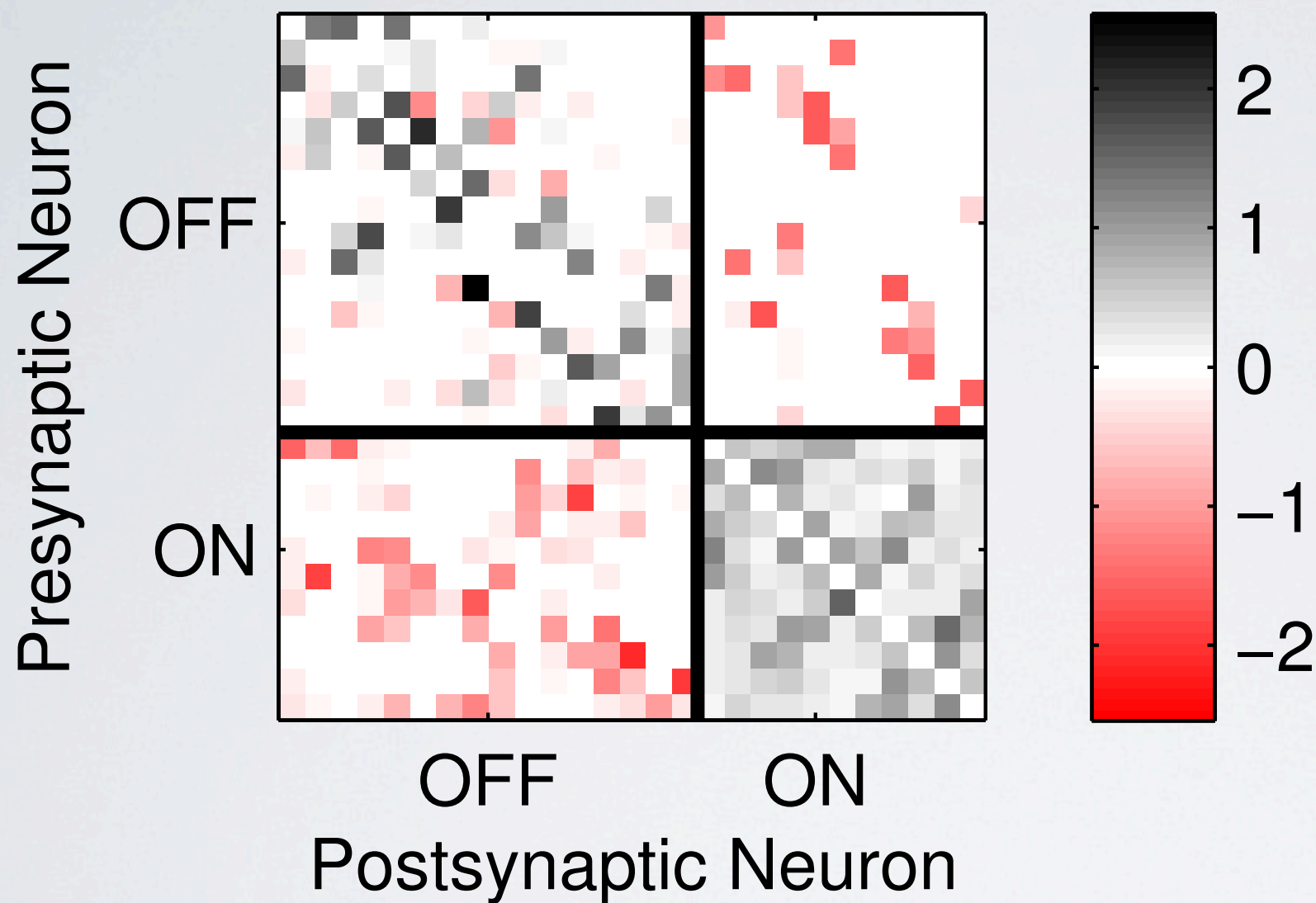
27 neurons, 50K spikes, macaque retina
(Data from Pillow and Chichilnisky)



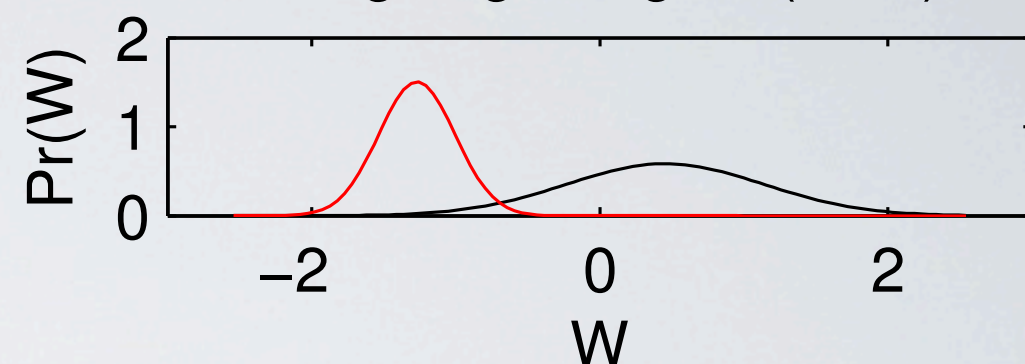


INFERRED NETWORK PROPERTIES

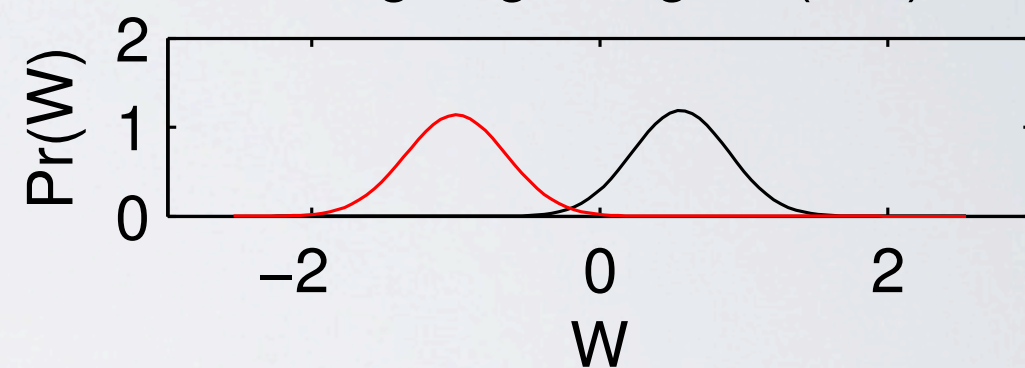
RGC Connectivity



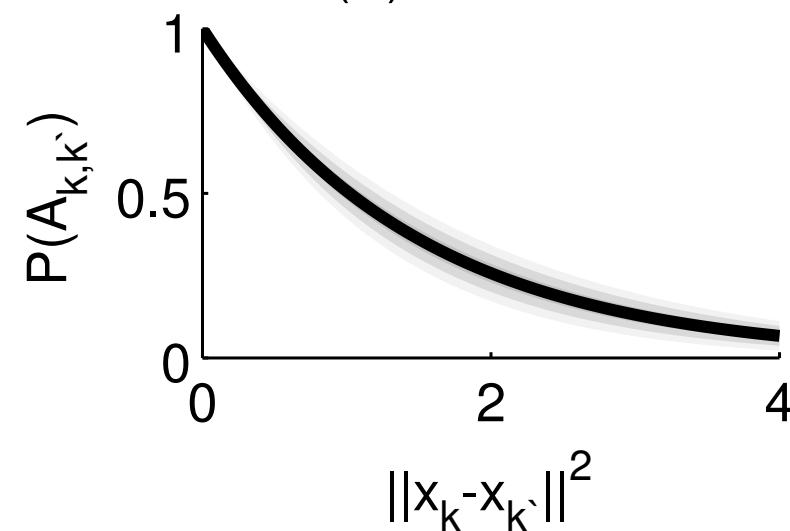
Outgoing Weights (OFF)

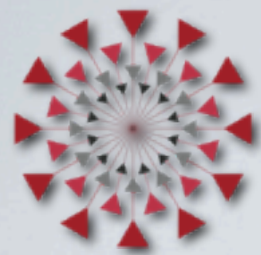


Outgoing Weights (ON)



Pr(A) vs Distance





SUMMARY

- ▶ We cannot directly observed edges for many networks of interest.
- ▶ However, these latent graphs can be inferred from vertex emissions.
- ▶ The purely-excitatory case (Hawkes process) enables an elegant data-augmentation approach to inference.
- ▶ MCMC is fast and tractable, and lets us reason about different graphs and graph priors.
- ▶ The inhibitory case is less tractable, but important for neuroscience applications.



Scott Linderman

